SAND2018-3989R

Compressed sensing with sparse corruptions: Fault-tolerant sparse collocation approximations

Ben Adcock*
Department of Mathematics
Simon Fraser University
Burnaby, BC, Canada

Anyi Bao*
Department of Mathematics
Simon Fraser University
Burnaby, BC, Canada

John D. Jakeman†
Computer Science Research Institute
Sandia National Laboratories
Albuquerque, NM, USA

Akil Narayan‡
Department of Mathematics and
Scientific Computing and Imaging (SCI) Institute
University of Utah
Salt Lake City, UT, USA

Abstract

The recovery of approximately sparse or compressible coefficients in a polynomial chaos expansion is a common goal in many modern parametric uncertainty quantification (UQ) problems. However, relatively little effort in UQ has been directed toward theoretical and computational strategies for addressing the sparse *corruptions* problem, where a small number of measurements are highly corrupted. Such a situation has become pertinent today since modern computational frameworks are sufficiently complex with many interdependent components that may introduce hardware and software failures, some of which can be difficult to detect and result in a highly polluted simulation result.

In this paper we present a novel compressive sampling-based theoretical analysis for a regularized ℓ^1 minimization algorithm that aims to recover sparse expansion coefficients in the presence of measurement corruptions. Our recovery results are uniform (the theoretical guarantees hold for all compressible signals and compressible corruptions vectors), and prescribe algorithmic regularization parameters in terms of a user-defined *a priori* estimate on the ratio of measurements that are believed to be corrupted. We also propose an iteratively reweighted optimization algorithm that automatically refines the value of the regularization parameter, and empirically produces superior results. Our numerical results test our framework on several medium-to-high dimensional examples of solutions to parameterized differential equations, and demonstrate the effectiveness of our approach.

*B. Adcock and A. Bao acknowledge the support of the Alfred P. Sloan Foundation and the Natural Sciences and Engineering Research Council of Canada through grant 611675.

†J.D.Jakeman's work was supported by DARPA EQUiPS.

‡A. Narayan is partially supported by NSF DMS-1552238, AFOSR FA9550-15-1-0467, and DARPA EQUiPS N660011524053

1 Introduction

The approximation of function values using point evaluations or samples is necessary in a wide number of applications. Much attention has been focused recently on the approximation technique of compressive sampling (CS): The ability to recover sparse linear representations of a function from a given dictionary. This is a particularly important problem in parametric uncertainty quantification (UQ) where the number of parameters translates into the number of variables on which an unknown function depends (the "dimension" of the problem). It is common for dimension to be very large, and the number of degrees of freedom in classical approximation strategies generally grows exponentially with the dimension. This makes classical computational procedures for approximating functions infeasible for large dimensions.

In contrast, compressive sampling seeks a sparse representation of a function using only a small number of samples or measurements, regardless of the parametric dimension. In a non-intrusive UQ pipeline, each function sample corresponds to a potentially large-scale simulation, and so minimizing the requisite number of samples is desirable. When functions are sparse or compressible in a given basis or dictionary, this reconstruction procedure has the potential to mitigate the exponentially debilitating curse of dimensionality. Algorithms in UQ that utilize compressive sampling have enjoyed great success in recent years [43, 42, 44, 27, 18, 10, 17, 15, 22]. For related theoretical contributions, see [1, 2, 8, 16, 29, 28, 41].

Missing from the sparse recovery UQ contributions above is a concrete strategy for fault-tolerant or resilient algorithms. Ensuring modeling resilience for UQ in the presence of system failures is essential for credible prediction on new and emerging massively parallel systems. Fault-tolerant algorithms in general have become necessary in computational science since node failures on distributed architectures can yield corrupted data (the frequency of which increases as the number of processors increases), or algorithmic run-time software failures can result in polluted simulation results. These failures can generate polluted measurements in unpredictable and sometimes undetectable ways [6].

Faults can occur due to complex combination of internal and external conditions that are difficult to reproduce. For example, bits may suffer random corruption, or physical defects in hardware may cause data faults. Corruption errors during model simulation can be grouped into two main types, soft and hard. In this paper, we consider hard faults as errors that cause the simulation to terminate prematurely and/or return obvious, automatically detectable error values such as NaN or Inf. Hard faults by this definition are easy to identify and mark for discard, thus obviating or ameliorating the need for fault-tolerant algorithms.

In contrast, soft failures are essentially random systematic corruption of results that are not easily identifiable. These soft failures pose challenges in UQ: A soft failure will not cause obvious failure in fault-intolerant UQ methods; however, incorrect model values caused by soft failures can significantly degrade an approximation. It is in this case that we require the development of robust and resilient algorithms that can, ideally, deliver constant levels of performance when faced with a few highly corrupted data points.

To address this issue, fault-tolerant algorithms for UQ have been investigated in the context of multilevel Monte Carlo algorithms [24, 25, 26], and in overdetermined least-squares polynomial recovery problems [31]. To the best of our knowledge, there is no comprehensive research in the UQ literature on fault-tolerant sparse recovery algorithms, and in the compressive sampling literature only a handful of papers [19, 21, 23, 32, 33, 34, 35, 39] deal with the problem of corrupted measurements.

The operative distinction in the problem we consider in this paper is a hardware or software fault resulting in occasional large-magnitude errors; we call this the problem of *corruptions*. Existing CS algorithms are known to be stable with respect to small noise perturbations, but cannot handle sparse corruptions, i.e., situations when a small number of samples are highly corrupted with the corruption magnitudes much larger than typical noise. In this paper we present novel theory and application studies of a sparse corruptions algorithm for CS. The algorithm we use was considered in [21], but we present more general theoretical guarantees on recovery, including practical guidance for the choice of algorithmic regularization parameters.

For fault-tolerance in the context of the sparse recovery problem, the recovery properties of an ideal resilient algorithm would be agnostic to large-magnitude corruptions in a small number of function samples. As described above, these corruptions can arise due to unknown failure modes in computational models or because of large but intermittent measurement errors. Development of mathematical theory for the corrupted compressive sampling problem, and investigation of a corresponding resilient algorithm for sparse

recovery of expansion coefficients are the central goals of this paper. The target applications we investigate are exemplars of a common task in UQ: recovery of approximately sparse expansion coefficients in an orthogonal polynomial (polynomial chaos) basis.

The theory and algorithms developed in this paper have the following features:

- The compressive sampling recovery theorems are uniform with respect to the function and the corruptions. That is, the recovery guarantees hold over all compressible functions having sparsely corrupted measurements for a single random sampling of measurements.
- The algorithm involves a tunable regularization parameter λ, and a theoretically optimal choice of this parameter is explicitly determined by our analytical results. This theoretically optimal value is defined only by the *ratio* of measurement corruptions to signal sparsity. Since signal sparsity is frequently comparable to the number of measurements, this optimal λ loosely translates into the fraction of measurement samples that are corrupted. From a user's point of view, our analysis thus suggests a value of λ having knowledge only of the ratio of measurements believed to be corrupted.
- In experiments, we observe that optimal values of the regularization parameter are non-trivially dependent on the number of measurements, the signal sparsity, and the number of corruptions. We thus propose an iteratively reweighted algorithm for recovery that learns values of the regularization parameter. Our experiments suggest that these learned algorithmic parameters perform better than the value defined by our theoretical results, and thus this reweighted algorithm is more useful in practice.
- The location and magnitude of the corruptions amongst the collection of function samples can be unknown, but the algorithm recovers those locations and the corresponding corruption values.
- The algorithm is robust to small, but non-sparse measurement errors – e.g. due to noise, truncation of an infinite polynomial expansion or numerical error in computing function samples – and moreover is *noise-blind*. That is to say, it requires no *a priori* upper bound on such errors.
- The optimization problem we solve to compute solutions is from [21], but our work is both a theoretical and practical advancement over the results in that reference. In order to show the solution computed is indeed the original sparse solution, [21] uses conditions on the restricted isometry constant (RIC) of the measurement matrix. Our results are a significant relaxation of previously reported conditions on the RIC (compare conditions on $\delta_{2s,2k}$ in Lemma 2.3 of [21] versus our Theorem 3.7, equation (3.8), and the discussion in Section 3.4). The results for general sensing matrices in [21] are nonuniform with respect to the signal and corruptions support, and require certain models for the signal and corruptions; our results are uniform and require no model for the signal or corruptions, other than compressibility. Finally, our paper is also devoted to numerical investigation of the performance of the method, including practical guidance for choosing the regularization parameter λ; such thorough investigations are absent in [21].

We first introduce notation and summarize the main mathematical statements of this paper in Section 2. This is followed in Section 3 by our theoretical analysis. Section 4 presents numerical results to complement our theoretical analysis and verify the practical efficacy of the algorithm.

2 Model problem and main results

Let $f : \mathbb{R}^d \to \mathbb{R}$ denote an unknown function, and let $\{\phi_j\}_{j=1}^N$ be a given dictionary of functions, $\phi_j : \mathbb{R}^d \to \mathbb{R}$. For example, the functions ϕ_j are frequently multivariate polynomial chaos basis elements; our capstone numerical examples will show results from such a basis. In scenarios of interest, the size N of the dictionary is very large.

The ultimate goal is to recover coefficients x_j that determine the approximation

$$f(\xi) = \sum_{j=1}^{N} x_j \phi_j(\xi) + n(\xi), \tag{2.1}$$

using samples of f, where $n(\xi)$ is an assumed small discrepancy term between the exact function and its N-term linear approximation in ϕ_j[1]. For the purposes of exposition we assume $|n(\xi)| \leq \epsilon$ for some known uniform noise bound ϵ; we will show later that lack of *a priori* knowledge for this bound only affects theoretical results in benign ways. As described above, we assume the vector $x = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$ to be compressible. Sparsity or compressibility of a vector can be quantified via its best s-term approximation error,

$$\sigma_s(x)_p = \inf_{\|\tilde{x}\|_0 \leq s} \|x - \tilde{x}\|_p ,$$

where $\|\cdot\|_p$ is the standard ℓ^p norm on vectors; for $p = 0$, $\|x\|_0$ is the sparsity of x, i.e., the number of non-zero elements in the vector.

With $\{\xi_1, \ldots, \xi_m\} \subset \mathbb{R}^d$ a collection of samples of ξ, we have the corresponding corrupted function measurements,

$$y_k = f(\xi_k) + c_k = \sum_{j=1}^{N} x_j \phi_j(\xi_k) + n(\xi_k) + c_k, \qquad k = 1, \ldots m,$$

where the corruption vector $c = (c_1, \ldots c_m)^T \in \mathbb{R}^m$ is assumed to be k-sparse but can have large entries. To enforce an underdetermined system, we assume $m < N$. Defining the rectangular matrix A with entries $(A)_{j,k} = \phi_k(\xi_j)$, then the unknown vectors x and c satisfy the underdetermined linear system

$$y = Ax + c + n \in \mathbb{R}^m. \tag{2.2}$$

In order to compute the solution (x, c) having knowledge of only A and y, we consider the following model problem (see also [21] and references therein):

$$\min_{z \in \mathbb{C}^N, d \in \mathbb{C}^m} \|z\|_1 + \lambda \|d\|_1 \text{ subject to } \|Az + d - y\|_2 \leq \epsilon\sqrt{m}. \tag{2.3}$$

Let $(\hat{x}, \hat{c})$ be a minimizer of this problem, where $\hat{x} \in \mathbb{R}^N$ and $\hat{c} \in \mathbb{R}^m$. Our objective is to obtain conditions on A (in particular, on the number of measurements m) and λ such that the error

$$\|\hat{x} - x\|_2 + \|c - \hat{c}\|_2$$

can be bounded by the best approximation numbers $\sigma_s(x)_1$ and $\sigma_k(c)_1$, and the noise magnitude ϵ.

2.1 Main results

In all that follows, the statement $a \lesssim b$ means $a \leq Cb$ for some universal constant C. Our first main result shows that stable and robust recovery of x and c is implied by a certain modification of the classical Restricted Isometry Property (RIP) which incorporates the sparse corruptions term (Definition 3.5). Specifically, Theorem 3.7 establishes that if the matrix A satisfies the RIP for the corruptions problem of order $(2s, 2k)$ (see Definition 3.5) with constant $\delta_{2s,2k}$ satisfying

$$\delta_{2s,2k} < \frac{1}{\sqrt{1 + \left(\frac{1}{2\sqrt{2}} + \sqrt{\eta}\right)^2}}, \qquad \eta = \frac{s + \lambda^2 k}{\min\{s, \lambda^2 k\}}, \tag{2.4}$$

then the following error bounds hold:

$$\|x - \hat{x}\|_1 + \lambda \|c - \hat{c}\|_1 \lesssim \sigma_s(x)_1 + \lambda \sigma_k(c)_1 + \epsilon\sqrt{s + \lambda^2 k}, \tag{2.5a}$$

$$\|x - \hat{x}\|_2 + \|c - \hat{c}\|_2 \lesssim \left(1 + \eta^{1/4}\right)\left(\frac{\sigma_s(x)_1}{\sqrt{s}} + \frac{\sigma_k(c)_1}{\sqrt{k}} + \epsilon\right). \tag{2.5b}$$

[1] Our notation suggests that $n = n(\xi)$ depends explicitly and deterministically on ξ; however, our theory encompasses the case when n is a stochastic variable or process, e.g., independent Gaussian random variable additive perturbations of the measurements.

Our second main result (Theorem 3.15) provides explicit conditions on m, s and k for (2.4) to hold for matrices of so-called bounded orthonormal system [12, Chpt. 12]. Specifically, suppose that $\{\phi_j\}_{j=1}^N$ is an $L^2_{\mathrm{d}\nu}(D)$-orthonormal system, where ν is a probability measure and $D \subset \mathbb{R}^d$ its support. Define

$$K \coloneqq \max_{j=1,\dots,N} \sup_{\xi \in D} |\phi_j(\xi)| < \infty,$$

and let $A = \{\phi_j(\xi_i)\}_{i,j=1}^{m,N}$ where $\xi_1, \dots, \xi_m$ are drawn i.i.d. according to ν. If

$$\begin{aligned} m &\gtrsim \delta^{-2} \cdot K^2 \cdot s \cdot \left(\log^3(2s) \cdot \log(2N) + \log \epsilon^{-1}\right), \\ m &\gtrsim \delta^{-2} \cdot K \cdot s \cdot k, \end{aligned} \tag{2.6}$$

then with probability at least $1 - \epsilon$, the restricted isometry constant $\delta_{2s,2k}$ of the scaled matrix $\frac{1}{\sqrt{m}} A$ satisfies $\delta_{2s,2k} \leq \delta$.

One can see from these estimates that optimizing η over values of λ yields a minimum value of $\eta = 2$ when $\lambda^2 = s/k$. Assuming $s \sim m$, this provides a concrete determination of the parameter λ for use in (2.3) having knowledge only of the ratio of corrupted measurements. We note in passing that we do not believe that the second condition in (2.6) is sharp in the dependence on the product $s \cdot k$. Improvement of this to a condition of the form

$$m \gtrsim \delta^{-2} \cdot K \cdot k, \tag{2.7}$$

is left as a topic for future work. Note that such a condition is known for Gaussian random matrices. Moreover, a nonuniform recovery result with the scaling (2.7) for exactly sparse coefficients x and corruptions c having random sign patterns was given in [21]. See Section 3.3 for further discussion.

It is common in compressed sensing to assume some *a priori* known noise bound ϵ based on the user's knowledge of measurement noise or truncation error. Although there are some results that circumvent this assumption [1, 2], they typically yield somewhat weaker recovery guarantees. However, in the context of the sparse corruptions theory presented above, such prior knowledge of ϵ is not necessary for stable recovery: The error introduced by an unknown noise ϵ can be passed into theoretical estimates as a penalty of size ϵ. To see this, note that if we define $c' \coloneqq \frac{1}{\sqrt{m}}(c + n)$, then the system $y = Ax + c + n$ can be written as $\frac{1}{\sqrt{m}} y = \frac{1}{\sqrt{m}} Ax + c'$. Solving (2.3) by setting $\epsilon = 0$ results in the $\epsilon = 0$ version of the estimate (2.5b) with c' replacing c. However, the normalized best k-term approximation error to c' appearing in (2.5b) is stable with respect to noise perturbations:

$$\frac{\sigma_k(c')_1}{\sqrt{k}} \leq \frac{1}{\sqrt{km}} \left(\sigma_k(c)_1 + \|n\|_1\right) \leq \frac{\sigma_k(c)_1}{\sqrt{km}} + \sqrt{\frac{m}{k}} \epsilon.$$

Here $\epsilon \geq \|n\|_\infty$ is any bound for the perturbation n in the uniform norm. Using (2.7), we see that $\sqrt{\frac{m}{k}} \epsilon \lesssim \epsilon$, which is on the same order as the estimate (2.5b) that uses *a priori* knowledge of ϵ. A similar argument holds for the bound (2.5a).

While our theoretical results are thus insensitive to ignorance about small noise levels, we caution that it is always a good idea to use such information in practical recovery algorithms if available, e.g. as the result of cross validation. See, for example, [10, 43, 17].

2.2 Remarks on numerical results

We postpone presenting numerical results until the end of this paper in Section 4. However, some remarks on our findings are pertinent here in the context of the previous section's theory. First, the optimal value of $\lambda^2 = s/k$ that is suggested by (2.4) does not appear to be the computationally optimal value of λ. That this fixed value of λ is not the best is not surprising since the bounds (2.5) are derived using some loose inequalities. However, such bounds can be useful in understanding qualitative trends. Results from our experimentation do suggest that large values of λ more reliably recover corruptions when s/k is large (see Figures 1 and 2). This general trend in numerical results is consistent with the behavior of η in (2.4) as a function of λ when s/k is large.

m	number of measurements
N	length of sparse vector
x	sparse vector in $\mathbb{C}^N$
c	corruptions vector in $\mathbb{C}^m$
A	$m \times N$ measurement matrix
n	noise vector in $\mathbb{C}^m$
ϵ	noise bound
λ	non-negative weighting parameter for the corruptions vector
$\hat{x}$, $\hat{c}$	solutions of the optimization problem
S	subset of $\{1, \ldots, N\}$, indices corresponding to x
T	subset of $\{1, \ldots, m\}$, indices corresponding to c
s	sparsity of x
k	sparsity of c
Σ_s	set of s-sparse vectors in $\mathbb{C}^N$
Σ_k	set of k-sparse vectors in $\mathbb{C}^m$
$\sigma_s(x)_1$	best s-term approximation error, measured in the ℓ^1 norm
$\sigma_k(c)_1$	best k-term approximation error, measured in the ℓ^1 norm

Table 1: Notation used throughout this article.

We address this discrepancy between the theory and empirical results by propose an iteratively reweighted ℓ^1 optimization scheme (see [7]) that learns and updates the value of λ. Our results show that this proposed algorithm performs much better in practice than algorithms that fix λ. However, we do not present any theory to support the observed superiority of reweighted ℓ^1 optimization schemes for the corruptions problem.

Many of our capstone numerical examples are from applications using polynomial chaos expansions, where the compressible function has an expansion in a multivariate orthogonal polynomial basis. To simplify the presentation of our results, we focus on such examples where the basis is a tensor-product Legendre polynomial or Chebyshev polynomial system. Much recent work has shown that randomly generating measurements using samples from standard distributions (e.g., the uniform distribution) can accurately and near-optimally recover orthogonal polynomial expansions from such basis sets [28, 17, 41]. Recovery in more general polynomial spaces has been investigated [16, 18, 15], but these methods usually rely on sophisticated sampling strategies and optimal sampling schemes are still an active area of research.

3 Theory for the sparse corruptions problem

We recall and summarize our notation for the sparse corruptions problem in Table 1. Our previous discussion was framed for real-valued signals x and measurements y, but we now generalize to the complex-valued setting. This adds generality with no additional mathematical difficulty.

We follow a familiar path for deriving conditions on m such that ℓ^1 optimization problems recover sparse solutions (see, for example, [12]). Section 3.1 defines an appropriate robust Null Space Property (NSP) for the matrix A in the sparse corruptions setting. Under this property, we show that the recovery estimates (2.5) hold. In order to construct matrices A that satisfy the robust NSP, Section 3.2 generalizes the concept of the Restricted Isometry Property (RIP) for matrices to the sparse corruptions setting. That section shows that matrices satisfying the RIP for the sparse corruptions problem also satisfy the robust NSP. Sections 3.3 and 3.3.2 show that if the dictionary elements ϕ_j form a bounded orthonormal system, then under the condition (2.6), the matrix A satisfies the RIP with high probability. Finally, using these various results, we discuss a theoretically-optimal choice for λ in Section 3.4.

3.1 The Robust Null Space Property for the sparse corruptions problem

The following two definitions are generalizations of robust null space properties (cf. [12, Definition 4.17] and [12, Definition 4.21], respectively), and prescribe classes of matrices whose kernels do not contain

sparse vectors.

Definition 3.1. *Let $1 \leq s \leq N$, $1 \leq k \leq m$ and $\lambda > 0$. A matrix $A \in \mathbb{C}^{m \times N}$ satisfies the ℓ^1-robust null space property of order (s,k) with weight λ if there exist constants $0 < \rho < 1$ and $\tau > 0$ such that*

$$\|x_S\|_1 + \lambda \|c_T\|_1 \leq \rho \left(\|x_{S^c}\|_1 + \lambda \|c_{T^c}\|_1 \right) + \tau \|Ax + c\|_2, \quad \forall x \in \mathbb{C}^N, \ c \in \mathbb{C}^m,$$

for all sets $S \subseteq \{1, \ldots, N\}$ and $T \subseteq \{1, \ldots, m\}$ with $|S| \leq s$ and $|T| \leq k$. Above, S^c is the complement of S in $\{1, \ldots, N\}$, and similarly for T^c.

Definition 3.2. *Let $1 \leq s \leq N$, $1 \leq k \leq m$ and $\lambda > 0$. A matrix $A \in \mathbb{C}^{m \times N}$ satisfies the ℓ^2-robust null space property of order (s,k) with weight λ if there exist constants $0 < \rho < 1$ and $\tau > 0$ such that*

$$\sqrt{\|x_S\|_2^2 + \|c_T\|_2^2} \leq \frac{\rho}{\sqrt{s + \lambda^2 k}} \left(\|x_{S^c}\|_1 + \lambda \|c_{T^c}\|_1 \right) + \tau \|Ax + c\|_2, \quad \forall x \in \mathbb{C}^N, \ c \in \mathbb{C}^m, \tag{3.1}$$

for all sets $S \subseteq \{1, \ldots, N\}$ and $T \subseteq \{1, \ldots, m\}$ with $|S| \leq s$ and $|T| \leq k$.

These definitions yield the following two results:

Lemma 3.3. *If $A \in \mathbb{C}^{m \times N}$ satisfies the ℓ^2-robust null space property of order (s,k) with weight $\lambda > 0$ and constants $0 < \rho < 1$, $\tau > 0$ then it satisfies the ℓ^1-robust null space property of order (s,k) with weight $\lambda > 0$ and constants ρ, $\tau\sqrt{s + \lambda^2 k}$.*

Proof. Observe that

$$\|x_S\|_1 + \lambda \|c_T\|_1 \leq \sqrt{s} \|x_S\|_2 + \lambda \sqrt{k} \|c_T\|_2 \leq \sqrt{s + \lambda^2 k} \sqrt{\|x_S\|_2^2 + \|c_T\|_2^2}.$$

We now use the definition of the ℓ^2-robust null space property. □

Theorem 3.4. *Let $1 \leq s \leq N$, $1 \leq k \leq m$ and $\lambda > 0$ and suppose that $A \in \mathbb{C}^{m \times N}$ satisfies the ℓ^2-robust null space property of order (s,k) with weight λ. Let $x \in \mathbb{C}^N$, $c \in \mathbb{C}^m$, $y \in \mathbb{C}^m$ and $\epsilon > 0$ be such that $\|Ax + c - y\|_2 \leq \epsilon$, and suppose that $(\hat{x}, \hat{c})$ is a minimizer of*

$$\min_{z \in \mathbb{C}^N, d \in \mathbb{C}^m} \|z\|_1 + \lambda \|d\|_1 \text{ subject to } \|Az + d - y\|_2 \leq \epsilon.$$

Then

$$\|x - \hat{x}\|_1 + \lambda \|c - \hat{c}\|_1 \leq C_1 \left(\sigma_s(x)_1 + \lambda \sigma_k(c)_1 \right) + C_2 \sqrt{s + \lambda^2 k} \epsilon, \tag{3.2}$$

and

$$\|x - \hat{x}\|_2 + \|c - \hat{c}\|_2 \leq C_3 \left(1 + \eta^{1/4} \right) \left(\frac{\sigma_s(x)_1}{\sqrt{s}} + \frac{\sigma_k(c)_1}{\sqrt{k}} \right) + C_4 \left(1 + \eta^{1/4} \right) \epsilon, \tag{3.3}$$

where the constants C_1, C_2, C_3, C_4 depend on ρ and τ only and η is given by

$$\eta = \eta_{s,k}(\lambda) = \frac{s + \lambda^2 k}{\min\{s, \lambda^2 k\}}. \tag{3.4}$$

Proof. We first prove (3.2). Lemma 3.3 implies that A satisfies the ℓ^1-robust null space property. Let $S \subseteq \{1, \ldots, N\}$, $|S| \leq s$ and $T \subseteq \{1, \ldots, m\}$, $|T| \leq k$ be such that $\|x_{S^c}\|_1 = \sigma_s(x)_1$ and $\|c_{T^c}\|_1 = \sigma_k(c)_1$. Then, if $v = x - \hat{x}$ and $e = c - \hat{c}$ we have

$$\begin{aligned} \|x\|_1 + \lambda \|c\|_1 + \|v_{S^c}\|_1 + \lambda \|e_{T^c}\|_1 &\leq 2\|x_{S^c}\|_1 + \|x_S\|_1 + \lambda \left(2\|c_{T^c}\|_1 + \|c_T\|_1 \right) + \|\hat{x}_{S^c}\|_1 + \lambda \|\hat{c}_{T^c}\|_1 \\ &\leq 2\|x_{S^c}\|_1 + \|v_S\|_1 + \|\hat{x}\|_1 + \lambda \left(2\|c_{T^c}\|_1 + \|e_T\|_1 + \|\hat{c}\|_1 \right). \end{aligned}$$

Rearranging now gives

$$\begin{aligned} \|v_{S^c}\|_1 + \lambda \|e_{T^c}\|_1 &\leq \left(2\|x_{S^c}\|_1 + \|v_S\|_1 \right) + \lambda \left(2\|c_{T^c}\|_1 + \|e_T\|_1 \right) \\ &\quad + \left(\|\hat{x}\|_1 + \lambda \|\hat{c}\|_1 \right) - \left(\|x\|_1 + \lambda \|c\|_1 \right) \\ &\leq 2 \left(\|x_{S^c}\|_1 + \lambda \|c_{T^c}\|_1 \right) + \left(\|v_S\|_1 + \lambda \|e_T\|_1 \right), \end{aligned}$$

where in the second inequality we note that $\|x\|_1 + \lambda\|c\|_1 \geq \|\hat{x}\|_1 + \lambda\|\hat{c}\|_1$ since (x, c) is feasible and $(\hat{x}, \hat{c})$ is a minimizer. The ℓ^1-robust null space property now implies that

$$\|v_{S^c}\|_1 + \lambda\|e_{T^c}\|_1 \leq \frac{2}{1-\rho}\left(\|x_{S^c}\|_1 + \lambda\|c_{T^c}\|_1\right) + \frac{\tau\sqrt{s+\lambda^2 k}}{1-\rho}\|Av+e\|_2,$$

and since $\|x_{S^c}\|_1 = \sigma_s(x)_1$, $\|c_{T^c}\|_1 = \sigma_k(c)_1$ and

$$\|Av + e\|_2 \leq \|A\hat{x} + \hat{c} - y\|_2 + \|Ax + c - y\|_2 \leq 2\epsilon, \tag{3.5}$$

we deduce that

$$\|v_{S^c}\|_1 + \lambda\|e_{T^c}\|_1 \leq \frac{2}{1-\rho}\left(\sigma_s(x)_1 + \lambda\sigma_k(c)_1\right) + \frac{2\tau}{1-\rho}\sqrt{s+\lambda^2 k}\epsilon. \tag{3.6}$$

Finally, to complete the proof of (3.2) we argue as follows:

$$\begin{aligned}
\|v\|_1 + \lambda\|e\|_1 &\leq \|v_S\|_1 + \lambda\|e_T\|_1 + \|v_{S^c}\|_1 + \lambda\|e_{T^c}\|_1 \\
&\leq (1+\rho)\left(\|v_{S^c}\|_1 + \lambda\|e_{T^c}\|_1\right) + \tau\sqrt{s+\lambda^2 k}\|Av+e\|_2 \\
&\leq 2\frac{1+\rho}{1-\rho}\left(\sigma_s(x)_1 + \lambda\sigma_k(c)_1\right) + \frac{4}{1-\rho}\tau\sqrt{s+\lambda^2 k}\epsilon.
\end{aligned}$$

Here, we use the ℓ^1-robust null space property in the second step, and (3.5) and (3.6) in the third step.

We now consider (3.3). Writing $v = x - \hat{x}$ and $e = c - \hat{c}$ as before, let S be the index of the largest s elements of v in absolute value and T be the index set of the largest k elements of e in absolute value. Define

$$\theta_v = \min_{i\in S}|v_i|, \quad \theta_e = \min_{j\in T}|e_j|, \qquad \theta = \max\{\theta_v, \theta_e/\lambda\}.$$

Then

$$\|v_{S^c}\|_2^2 + \|e_{T^c}\|_2^2 = \sum_{i\notin S}|v_i|^2 + \sum_{j\notin T}|e_j|^2 \leq \theta_v\sum_{i\notin S}|v_i| + \theta_e\sum_{j\notin T}|e_j| \leq \theta\left(\|v_{S^c}\|_1 + \lambda\|e_{T^c}\|_1\right).$$

Now observe that $\theta_v \leq \|v_S\|_2/\sqrt{s}$ and $\theta_e \leq \|e_T\|_2/\sqrt{k}$, and therefore

$$\theta \leq \frac{\sqrt{\|v_S\|_2^2 + \|e_T\|_2^2}}{\min\{\sqrt{s}, \lambda\sqrt{k}\}} \leq \frac{1}{\min\{\sqrt{s}, \lambda\sqrt{k}\}}\left(\frac{\rho}{\sqrt{s+\lambda^2 k}}\left(\|v_{S^c}\|_1 + \lambda\|e_{T^c}\|_1\right) + 2\tau\epsilon\right),$$

where in the second step we use the ℓ^2-robust null space property and (3.5). Combining this with the previous estimate and using the definition of η gives

$$\begin{aligned}
\|v_{S^c}\|_2^2 + \|e_{T^c}\|_2^2 \leq& \frac{1}{\min\{\sqrt{s}, \lambda\sqrt{k}\}}\left(\frac{\rho}{\sqrt{s+\lambda^2 k}}\left(\|v_{S^c}\|_1 + \lambda\|e_{T^c}\|_1\right)^2 + 2\tau\epsilon\left(\|v_{S^c}\|_1 + \lambda\|e_{T^c}\|_1\right)\right) \\
=& \sqrt{\eta}\left[\rho w^2 + 2\tau\epsilon w\right],
\end{aligned}$$

where we have defined the non-negative scalar w as

$$w := \frac{\|v_{S^c}\|_1 + \lambda\|e_{T^c}\|_1}{\sqrt{s+\lambda^2 k}}$$

Completing the square with respect to w under the brackets yields

$$\|v_{S^c}\|_2^2 + \|e_{T^c}\|_2^2 \leq \rho\sqrt{\eta}\left[\left(w + \frac{\tau\epsilon}{\sqrt{\rho}}\right)^2 - \frac{\tau^2\epsilon^2}{\rho}\right] \leq \rho\sqrt{\eta}\left(w + \frac{\tau\epsilon}{\sqrt{\rho}}\right)^2$$

Using the ℓ^2-robust NSP on the pair (v, e) along with the above estimate, we have

$$\begin{aligned}\frac{1}{\sqrt{2}}\left(\|v\|_2+\|e\|_2\right) \leq \sqrt{\|v\|_2^2+\|e\|_2^2} &= \sqrt{\|v_S\|_2^2+\|e_T\|_2^2+\|v_{S^c}\|_2^2+\|e_{T^c}\|_2^2} \\ &\leq \sqrt{\|v_S\|_2^2+\|e_T\|_2^2}+\sqrt{\|v_{S^c}\|_2^2+\|e_{T^c}\|_2^2} \\ &\leq \rho w+2\tau\epsilon+\sqrt{\rho}\eta^{1/4}\left(w+\frac{\tau\epsilon}{\sqrt{\rho}}\right) \\ &= \sqrt{\rho}\left(\sqrt{\rho}+\eta^{1/4}\right)w+\tau\left(2+\eta^{1/4}\right)\epsilon \end{aligned} \tag{3.7}$$

We note that

$$\begin{aligned} w &= \frac{\|v_{S^c}\|_1+\lambda\|e_{T^c}\|_1}{\sqrt{s+\lambda^2 k}} \leq \frac{\|v\|_1+\lambda\|e\|_1}{\sqrt{s+\lambda^2 k}} \\ &\overset{(3.2)}{\leq} C_1\left[\frac{\sigma_s(x)_1}{\sqrt{s+\lambda^2 k}}+\lambda\frac{\sigma_k(c)_1}{\sqrt{s+\lambda^2 k}}\right]+C_2\epsilon \leq C_1\left[\frac{\sigma_s(x)_1}{\sqrt{s}}+\frac{\sigma_k(c)_1}{\sqrt{k}}\right]+C_2\epsilon \end{aligned}$$

Combining the above with (3.7) proves (3.3). □

3.2 The Restricted Isometry Property for the sparse corruptions problem

The robust NSP is typically difficult to prove directly. Hence we now introduce the Restricted Isometry Property (RIP) for the sparse corruptions problem, and show that it implies the robust NSP. Note that this has been defined previously in [21, Defn. 2.1].

Definition 3.5. *Let $1 \leq s \leq N$, $1 \leq k \leq m \leq N$ and $A \in \mathbb{C}^{m\times N}$. The $(s,k)^{\text{th}}$ Restricted Isometry Constant (RIC) $\delta = \delta_{s,k}$ of the matrix A is the smallest constant such that*

$$(1-\delta)\left(\|x\|_2^2+\|c\|_2^2\right) \leq \|Ax+c\|_2^2 \leq (1+\delta)\left(\|x\|_2^2+\|c\|_2^2\right)$$

for all $x \in \Sigma_s$ and $c \in \Sigma_k$. If $0 < \delta_{s,k} < 1$ then we say that A has the Restricted Isometry Property (RIP) of order (s,k).

Our first result is the following:

Lemma 3.6. *Let $1 \leq s \leq N$, $1 \leq k \leq m \leq N$, $\lambda > 0$ and $A \in \mathbb{C}^{m\times N}$. If A satisfies the RIP of order $(2s, 2k)$ with constant*

$$\delta_{2s,2k} < \frac{1}{\sqrt{1+\left(\frac{1}{2\sqrt{2}}+\sqrt{\eta}\right)^2}}, \tag{3.8}$$

where η is as in (3.4), then A satisfies the ℓ^2-robust NSP of order (s,k) with weight λ and constants $0 < \rho < 1$ and $\tau > 0$ depending only on $\delta_{2s,2k}$.

The proof of this result is given next. Combining this lemma with Theorem 3.4 now yields our main result:

Theorem 3.7. *Let $1 \leq s \leq N$, $1 \leq k \leq m$ and $\lambda > 0$ and suppose that $A \in \mathbb{C}^{m\times N}$ satisfies the RIP of order $(2s, 2k)$ with constant $\delta_{2s,2k}$ satisfying (3.8) and η as in (3.4). Let $x \in \mathbb{C}^N$, $c \in \mathbb{C}^m$, $y \in \mathbb{C}^m$ and $\epsilon > 0$ be such that $\|Ax+c-y\|_2 \leq \epsilon$, and suppose that $(\hat{x}, \hat{c})$ is a minimizer of*

$$\min_{z\in\mathbb{C}^N, d\in\mathbb{C}^m} \|z\|_1+\lambda\|d\|_1 \text{ subject to } \|Az+d-y\|_2 \leq \epsilon,$$

Then

$$\begin{aligned} \|x-\hat{x}\|_1+\lambda\|c-\hat{c}\|_1 &\leq C_1\left(\sigma_s(x)_1+\lambda\sigma_k(c)_1\right)+C_2\sqrt{s+\lambda^2 k}\epsilon, \\ \|x-\hat{x}\|_2+\|c-\hat{c}\|_2 &\leq C_3\left(1+\eta^{1/4}\right)\left(\frac{\sigma_s(x)_1}{\sqrt{s}}+\frac{\sigma_k(c)_1}{\sqrt{k}}\right)+C_4\left(1+\eta^{1/4}\right)\epsilon, \end{aligned}$$

where the constants C_1, C_2, C_3, C_4 depend on $\delta_{2s,2k}$ only.

We now prove Lemma 3.6. We first require the following:

Lemma 3.8. *Let $1 \leq s \leq N$, $1 \leq k \leq m \leq N$, and let $A \in \mathbb{C}^{m \times N}$ satisfy the RIP of order $(2s, 2k)$ with constant $\delta_{2s,2k}$. Suppose that $x \in \Sigma_s$ and $c \in \Sigma_k$ are such that*

$$\|Ax + c\|_2^2 - \left(\|x\|_2^2 + \|c\|_2^2\right) = t\left(\|x\|_2^2 + \|c\|_2^2\right),$$

for some t with $0 \leq |t| \leq \delta_{2s,2k}$. If $z \in \Sigma_s$ and $d \in \Sigma_k$ are orthogonal to x and c, respectively, then

$$|\langle Ax + c, Az + d\rangle| \leq \sqrt{\delta_{2s,2k}^2 - t^2}\sqrt{\|x\|_2^2 + \|c\|_2^2}\sqrt{\|z\|_2^2 + \|d\|_2^2}.$$

Proof. Assume that $\|x\|_2^2 + \|c\|_2^2 = \|z\|_2^2 + \|d\|_2^2 = 1$ without loss of generality. Let $\alpha, \beta \in \mathbb{R}$ and $\gamma \in \mathbb{C}$ and notice that $\alpha x + \gamma z, \beta x - \gamma z \in \Sigma_{2s}$ and $\alpha c + \gamma d, \beta c - \gamma d \in \Sigma_{2k}$. Therefore

$$\begin{aligned}
\|A(\alpha x + \gamma z) + (\alpha c + \gamma d)\|_2^2 &\leq (1 + \delta_{2s,2k})\left(\|\alpha x + \gamma z\|_2^2 + \|\alpha c + \gamma d\|_2^2\right) \\
&= (1 + \delta_{2s,2k})\left(\alpha^2\left(\|x\|_2^2 + \|c\|_2^2\right) + |\gamma|^2\left(\|z\|_2^2 + \|d\|_2^2\right)\right) \\
&= (1 + \delta_{2s,2k})\left(\alpha^2 + |\gamma|^2\right).
\end{aligned}$$

Note that in the second step we use orthogonality of the vectors x and z and c and d. Similarly,

$$\|A(\beta x - \gamma z) + (\beta c - \gamma d)\|_2^2 \geq (1 - \delta_{2s,2k})\left(\beta^2 + |\gamma|^2\right).$$

Subtracting the second equation from the first gives

$$\begin{aligned}
\|A(\alpha x + \gamma z) + (\alpha c + \gamma d)\|_2^2 - &\|A(\beta x - \gamma z) + (\beta c - \gamma d)\|_2^2 \\
&\leq (1 + \delta_{2s,2k})\left(\alpha^2 + |\gamma|^2\right) - (1 - \delta_{2s,2k})\left(\beta^2 + |\gamma|^2\right) \\
&= \delta_{2s,2k}\left(\alpha^2 + \beta^2 + 2|\gamma|^2\right) + \alpha^2 - \beta^2. \qquad (3.9)
\end{aligned}$$

On the other hand

$$\begin{aligned}
\|A(\alpha x + \gamma z) + (\alpha c + \gamma d)\|_2^2 - &\|A(\beta x - \gamma z) + (\beta c - \gamma d)\|_2^2 \\
&= \alpha^2\|Ax + c\|_2^2 + |\gamma|^2\|Az + d\|_2^2 + 2\mathrm{Re}\,\langle\alpha(Ax + c), \gamma(Az + d)\rangle \\
&\quad - \beta^2\|Ax + c\|_2^2 - |\gamma|^2\|Az + d\|_2^2 + 2\mathrm{Re}\,\langle\beta(Ax + c), \gamma(Az + d)\rangle \\
&= \left(\alpha^2 - \beta^2\right)\|Ax + c\|_2^2 + 2(\alpha + \beta)\mathrm{Re}\,\left(\bar{\gamma}\langle Ax + c, Az + d\rangle\right) \\
&= \left(\alpha^2 - \beta^2\right)(1 + t) + 2(\alpha + \beta)\mathrm{Re}\,\left(\bar{\gamma}\langle Ax + c, Az + d\rangle\right).
\end{aligned}$$

Combining this with (3.9) gives

$$\left(\alpha^2 - \beta^2\right)(1 + t) + 2(\alpha + \beta)\mathrm{Re}\,\left(\bar{\gamma}\langle Ax + c, Az + d\rangle\right) \leq \delta_{2s,2k}\left(\alpha^2 + \beta^2 + 2|\gamma|^2\right) + \alpha^2 - \beta^2.$$

Now let γ be such that $|\gamma| = 1$ and $\mathrm{Re}\,\left(\bar{\gamma}\langle Ax + c, Az + d\rangle\right) = |\langle Ax + c, Az + d\rangle|$. Then, after rearranging, we get

$$|\langle Ax + c, Az + d\rangle| \leq \frac{\left(\delta_{2s,2k} - t\right)\alpha^2 + \left(\delta_{2s,2k} + t\right)\beta^2 + 2\delta_{2s,2k}}{2(\alpha + \beta)}.$$

We now seek values α and β which minimize the right-hand side of this expression. If $t = \delta_{2s,2k}$ then the minimal value 0 is attained by setting $\beta = 0$ and letting $\alpha \to \infty$. Conversely, if $t < \delta_{2s,2k}$ the minimal value is attained when $\alpha = \sqrt{\frac{\delta_{2s,2k}+t}{\delta_{2s,2k}-t}}$ and $\beta = \frac{1}{\alpha}$. This gives

$$|\langle Ax + c, Az + d\rangle| \leq \sqrt{\delta_{2s,2k}^2 - t^2},$$

which completes the proof. □

Proof of Lemma 3.6. Let $x \in \mathbb{C}^N$ and $c \in \mathbb{C}^m$. To prove the ℓ^2-robust NSP for A it is enough to show that (3.1) holds when $S = S_0$ is the index set of the s largest coefficients of x in absolute value and $T = T_0$ is the set of the k largest values of c in absolute value. Given S_0, let S_1 be the index set of the next s largest coefficients of x in absolute value, S_2 be the index set of the next s largest coefficients and so on. Define $T_1, T_2, \dots$ in a similar way. We now have the following:

$$\begin{aligned}\|Ax_{S_0} + c_{T_0}\|^2 &= \langle Ax_{S_0} + c_{T_0}, Ax_{S_0} + c_{T_0}\rangle \\ &= \langle Ax_{S_0} + c_{T_0}, Ax + c\rangle - \sum_{j\geq 1}\langle Ax_{S_0} + c_{T_0}, Ax_{S_j} + c_{T_j}\rangle. \end{aligned} \tag{3.10}$$

Let $0 \leq |t| \leq \delta_{2s,2k}$ be such that

$$\|Ax_{S_0} + c_{T_0}\|_2^2 = (1+t)\left(\|x_{S_0}\|_2^2 + \|c_{T_0}\|_2^2\right), \tag{3.11}$$

and note that this gives

$$|\langle Ax_{S_0} + c_{T_0}, Ax + c\rangle| \leq \sqrt{1+t}\sqrt{\|x_{S_0}\|_2^2 + \|c_{T_0}\|_2^2}\|Ax + c\|_2. \tag{3.12}$$

For the second term of (3.10), we use the disjointness of S_0 and S_j and T_0 and T_j for $j \geq 1$ in combination with Lemma 3.8 to get

$$\begin{aligned}\left|\sum_{j\geq 1}\langle Ax_{S_0} + c_{T_0}, Ax_{S_j} + c_{T_j}\rangle\right| &\leq \sqrt{\delta_{2s,2k}^2 - t^2}\sqrt{\|x_{S_0}\|_2^2 + \|c_{T_0}\|_2^2}\sum_{j\geq 1}\sqrt{\|x_{S_j}\|_2^2 + \|c_{T_j}\|_2^2} \\ &\leq \sqrt{\delta_{2s,2k}^2 - t^2}\sqrt{\|x_{S_0}\|_2^2 + \|c_{T_0}\|_2^2}\left(\sum_{j\geq 1}\|x_{S_j}\|_2 + \sum_{j\geq 1}\|c_{T_j}\|_2\right). \end{aligned} \tag{3.13}$$

Let x_j^+ and x_j^- be the largest entries of x_{S_j} in absolute value. Then, by [12, Lem. 6.14], we have

$$\begin{aligned}\sum_{j\geq 1}\|x_{S_j}\|_2 &\leq \sum_{j\geq 1}\left(\frac{\|x_{S_j}\|_1}{\sqrt{s}} + \frac{\sqrt{s}}{4}\left(x_j^+ - x_j^-\right)\right) \\ &\leq \frac{\|x_{S_0^c}\|_1}{\sqrt{s}} + \frac{\sqrt{s}}{4}\sum_{j\geq 1}\left(x_j^+ - x_{j+1}^+\right) \leq \frac{\|x_{S_0^c}\|_1}{\sqrt{s}} + \frac{\sqrt{s}}{4}x_1^+ \leq \frac{\|x_{S_0^c}\|_1}{\sqrt{s}} + \frac{1}{4}\|x_{S_0}\|_2. \end{aligned}$$

Similarly,

$$\sum_{j\geq 1}\|c_{T_j}\|_2 \leq \frac{\|c_{T_0^c}\|_1}{\sqrt{k}} + \frac{1}{4}\|c_{T_0}\|_2 \leq \frac{\lambda\|c_{T_0^c}\|_1}{\lambda\sqrt{k}} + \frac{1}{4}\|c_{T_0}\|_2,$$

which gives

$$\sum_{j\geq 1}\|x_{S_j}\|_2 + \sum_{j\geq 1}\|c_{T_j}\|_2 \leq \frac{1}{\min\left\{\sqrt{s}, \lambda\sqrt{k}\right\}}\left(\|x_{S_0^c}\|_1 + \lambda\|c_{T_0^c}\|_1\right) + \frac{1}{4}\left(\|x_{S_0}\|_2 + \|c_{T_0}\|_2\right).$$

Therefore, combining this with (3.10), (3.11), (3.12) and (3.13) yields

$$\begin{aligned}(1+t)\sqrt{\|x_{S_0}\|_2^2 + \|c_{T_0}\|_2^2} &\leq \sqrt{1+t}\|Ax + c\|_2 \\ &+ \sqrt{\delta_{2s,2k}^2 - t^2}\left(\frac{1}{\min\left\{\sqrt{s}, \lambda\sqrt{k}\right\}}\left(\|x_{S_0^c}\|_1 + \lambda\|c_{T_0^c}\|_1\right) + \frac{1}{4}\left(\|x_{S_0}\|_2 + \|c_{T_0}\|_2\right)\right). \end{aligned}$$

Consider the function $g(t) = \frac{\delta_{2s,2k}^2 - t^2}{(1+t)^2}$, where $0 \leq t \leq \delta_{2s,2k}$. This function attains its maximum value at $t = -\delta_{2s,2k}^2$ and takes value $\frac{\delta_{2s,2k}^2}{1-\delta_{2s,2k}^2}$ there. Additionally $\frac{1}{\sqrt{1+t}} \leq \frac{1}{\sqrt{1-\delta_{2s,2k}}}$. Hence we get

$$\begin{aligned}\sqrt{\|x_{S_0}\|_2^2 + \|c_{T_0}\|_2^2} \leq{}& \frac{1}{\sqrt{1-\delta_{2s,2k}}}\|Ax+c\|_2 \\ &+ \frac{\delta_{2s,2k}}{\sqrt{1-\delta_{2s,2k}^2}}\left(\frac{1}{\min\left\{\sqrt{s}, \lambda\sqrt{k}\right\}}\left(\left\|x_{S_0^c}\right\|_1 + \left\|c_{T_0^c}\right\|_1\right) + \frac{1}{4}\left(\|x_{S_0}\|_2 + \|c_{T_0}\|_2\right)\right).\end{aligned}$$

After noting that $\|x_{S_0}\|_2 + \|c_{T_0}\|_2 \leq \sqrt{2}\sqrt{\|x_{S_0}\|_2^2 + \|c_{T_0}\|_2^2}$ and rearranging, we obtain

$$\sqrt{\|x_{S_0}\|_2^2 + \|c_{T_0}\|_2^2} \leq \frac{\rho}{\sqrt{s+\lambda^2 k}}\left(\left\|x_{S_0^c}\right\|_1 + \left\|c_{T_0^c}\right\|_1\right) + \tau\|Ax+c\|_2,$$

where

$$\rho = \frac{2\sqrt{2}\delta_{2s,2k}}{2\sqrt{2}\sqrt{1-\delta_{2s,2k}^2} - \delta_{2s,2k}}\sqrt{\eta}, \quad \tau = \frac{2\sqrt{2}\sqrt{1+\delta_{2s,2k}}}{2\sqrt{2}\sqrt{1-\delta_{2s,2k}^2} - \delta_{2s,2k}}. \tag{3.14}$$

To complete the proof we note that $\rho, \tau > 0$ provided $\delta_{2s,2k} < \sqrt{8/9}$. This holds by assumption, since $\eta \geq 2$ and therefore the condition (3.8) implies that $\delta_{2s,2k} < \sqrt{8/33} < \sqrt{8/9}$. Also, after rearranging we see that $\rho < 1$ if

$$\left(1 + \left(\frac{1}{2\sqrt{2}} + \sqrt{\eta}\right)^2\right)\delta_{2s,2k}^2 < 1,$$

which again holds by assumption. □

Remark 3.9 The RIP for the sparse corruptions problem is a special case of the RIP in levels (RIPL), introduced in [5]. The RIPL applies to vectors that are sparse in levels; namely, having different amounts of sparsity in different (but fixed) sections of the vector. In the context of the sparse corruptions problem, this corresponds to the concatenated vector $z = [x; c]$, which is s-sparse in its first N entries and k-sparse in its remaining m entries. As a general tool, sparsity in levels has been used in the context of compressive imaging [3, 4, 30], radar [11] and multi-sensor acquisition [9]. It is interesting that the same model also occurs naturally in the, seemingly unrelated, sparse corruptions problem. We note in passing that Theorems 3.4 and 3.7 follow a similar approach to that of [5] with some changes made to incorporate the weighted optimization problem.

3.3 Matrices that satisfy the RIP for sparse corruptions

We first recall the classical RIP for sparse vectors:

Definition 3.10. *Let $1 \leq s \leq N$ and $A \in \mathbb{C}^{m\times N}$. The s^{th} Restricted Isometry Constant (RIC) $\delta = \delta_s$ of the matrix A is the smallest constant such that*

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2,$$

for all $x \in \Sigma_s$. If $0 < \delta_s < 1$ then we say that A has the Restricted Isometry Property (RIP) of order s.

To distinguish it from the RIP for the sparse corruptions problem (Definition 3.5), we shall refer to this as the *RIP for sparse vectors*.

Lemma 3.11. *Let $1 \leq s \leq N$, $1 \leq k \leq m$, $A \in \mathbb{C}^{m\times N}$ and define*

$$\sigma_{s,k} = \max_{\substack{S\subseteq\{1,\dots,N\},|S|=s \\ T\subseteq\{1,\dots,m\},|T|=k}} \|A_{S,T}\|_2, \tag{3.15}$$

where $A_{S,T} \in \mathbb{C}^{|T|\times|S|}$ is the submatrix of A with entries $\{A_{ij}\}_{i\in T, j\in S}$. Suppose that A has the RIP for sparse vectors with constant δ_s and that $\sigma_{s,k} < \sqrt{1-\delta_s}$. Then A has the RIP of order (s,k) for the sparse corruptions problem with constant

$$\delta_{s,k} = \frac{\delta_s + \sqrt{\delta_s^2 + 4\sigma_{s,k}^2}}{2}.$$

In other words,

$$(1-\delta_{s,k})\left(\|x\|_2^2 + \|c\|_2^2\right) \leq \|Ax+c\|_2^2 \leq (1+\delta_{s,k})\left(\|x\|_2^2 + \|c\|_2^2\right)$$

for all $x \in \Sigma_s$ and $c \in \Sigma_k$.

Proof. Let $x \in \Sigma_s$ and $c \in \Sigma_k$ and write $S = \mathrm{supp}(x)$ and $T = \mathrm{supp}(c)$. Then

$$\|Ax+c\|_2^2 = \|Ax\|_2^2 + \|c\|_2^2 + 2\mathrm{Re}\,\langle A_{S,T}x, c\rangle.$$

By Young's inequality

$$2\,|\langle A_{S,T}x, c\rangle| \leq 2\|A_{S,T}\|_2\|x\|_2\|c\|_2 \leq \|A_{S,T}\|_2\left(\|x\|_2^2/\epsilon + \epsilon\|c\|_2^2\right),$$

for any $\epsilon > 0$. Hence

$$(1-\delta_s-\sigma_{s,k}/\epsilon)\,\|x\|_2^2 + (1-\sigma\epsilon)\,\|c\|_2^2 \leq \|Ax+c\|_2^2 \leq (1+\delta_s+\sigma_{s,k}/\epsilon)\,\|x\|_2^2 + (1+\sigma\epsilon)\,\|c\|_2^2.$$

Solving the equation $\delta_s + \sigma_{s,k}/\epsilon = \sigma_{s,k}\epsilon$ yields the value $\epsilon = \frac{\delta_s + \sqrt{\delta_s^2+4\sigma_{s,k}^2}}{2\sigma}$, and substituting this value of ϵ into the previous expression yields the proof. □

This result shows that any matrix satisfying the RIP for sparse vectors also satisfies the RIP for the sparse corruptions problem, provided the all $k \times s$ submatrices have small spectral norm.

3.3.1 Gaussian random matrices

Gaussian random matrices in the context of the sparse corruptions problem were considered in [21]. The following result essentially recaps the main result for this case given therein. We include a short proof for completeness:

Theorem 3.12. *Let $0 < \delta, \epsilon < 1$, $1 \leq s \leq$, $1 \leq k \leq m$ and suppose that*

$$m \gtrsim \delta^{-2}\left(s \cdot \log(2N/s) + \log(2\epsilon^{-1})\right), \tag{3.16}$$

$$m \gtrsim \delta^{-2} \cdot k \cdot \log(\delta^{-1}). \tag{3.17}$$

Let $A \in \mathbb{C}^{m\times N}$ be a matrix whose entries are independent Gaussian random variables with mean zero and variance 1. Then with probability at least $1-\epsilon$, the matrix $\frac{1}{\sqrt{m}}A$ has the RIP for the sparse corruptions problem of order (s,k) with constant $\delta_{s,k} \leq \delta$.

Proof. Lemma 3.11 asserts that A has the RIP of order (s,k) for the sparse corruptions problem with constant $\delta_{s,k} \leq \delta$ provided (i) A has the RIP of order s with $\delta_s \leq \delta/\sqrt{2}$ and (ii) the constant $\sigma_{s,k}$ defined in (3.15) satisfies $\sigma_{s,k} \leq \delta/(2\sqrt{2})$. Hence, by the union bound it suffices to show that (3.16) and (3.17) imply both (i) and (ii) separately with probabilities at least $1-\epsilon/2$. Due to a standard result in compressed sensing (see, for example, [12, Thm. 9.2]), property (i) holds with probability at least $1-\epsilon/2$ whenever the condition (3.16) is satisfied. We now consider property (ii). First, notice that $\sigma_{s,k}$ is increasing in k. Therefore, we may assume that $k \asymp \delta^2 \cdot m$, i.e. $k \gtrsim \delta^2 \cdot m$ and $k \lesssim \delta^2 \cdot m$. Fix subsets $S \subseteq \{1,\ldots,N\}$ and $T \subseteq \{1,\ldots,m\}$ with $|S| = s$ and $|T| = k$. Then, due to a known result for singular values of random Gaussian matrices (see, for example, [38, Cor. 5.35]), we have

$$\mathbb{P}\left(\|A_{S,T}\|_2 \geq \sqrt{s} + \sqrt{k} + t\right) \leq 2\exp(-t^2/2).$$

The conditions (3.16) and (3.17) imply that $\sqrt{s/m} \leq \delta/(6\sqrt{2})$ and $\sqrt{k/m} \leq \delta/(6\sqrt{2})$. Hence, by the union bound

$$\mathbb{P}\left(\sigma_{s,k} > \delta/(2\sqrt{2})\right) \leq \left(\begin{array}{c} N \\ s \end{array}\right)\left(\begin{array}{c} m \\ k \end{array}\right) \exp(-m\delta^2/48) \leq \left(\frac{\mathrm{e}N}{s}\right)^s \left(\frac{\mathrm{e}m}{k}\right)^k \exp(-m\delta^2/48).$$

In particular, $\mathbb{P}\left(\sigma_{s,k} > \delta/(2\sqrt{2})\right) \leq \epsilon/2$ provided

$$m \geq 48 \cdot \delta^{-2}\left(s\log(\mathrm{e}N/s) + k\log(\mathrm{e}m/k) + \log(2\epsilon^{-1})\right).$$

Since $k \asymp \delta^2 \cdot m$, we have $\log(\mathrm{e}m/k) \lesssim \log(2\delta^{-1})$. Hence this condition is implied by (3.16) and (3.17). This establishes property (ii) and completes the proof. □

This result asserts that Gaussian random matrices can recover a fixed fraction $k/m \leq c$ of corruptions (see (3.17)) and (up to constants) the same level of sparsity s as in the uncorrupted case (see (3.16)).

3.3.2 Bounded orthonormal systems

Gaussian random matrices, while mathematically appealing, are of little relevance to multivariate approximation using Polynomial Chaos expansions. In this case, a more suitable framework is that of bounded orthonormal systems (see, for example, [12, Chpt. 12]):

Let D be a domain with a probability measure ν and $\phi_1, \ldots, \phi_N$ be an orthonormal system of complex-value functions in $L^2(D)$. Recall that this system is bounded if

$$\|\phi_i\|_{L^\infty} = \sup_{\xi \in D} |\phi_i(\xi)| \leq K$$

Given such a system, we construct the measurement matrix A as

$$A = \frac{1}{\sqrt{m}} \{\phi_j(\xi_i)\}_{i=1,j=1}^{m,N} \in \mathbb{C}^{m \times N}, \tag{3.18}$$

where t_i are drawn independently at random according to the probability measure ν.

Theorem 3.13. *Let $A \in \mathbb{C}^{m\times N}$ be the matrix of a bounded orthonormal system, $1 \leq s \leq N$ and $0 < \delta, \epsilon < 1$. If*

$$m \gtrsim \delta^{-2} \cdot s \cdot \left(\log^3(2s) \cdot \log(2N) + \log(\epsilon^{-1})\right),$$

then A satisfies the RIP for sparse vectors with probability at least $1 - \epsilon$.

We remark in passing that the logarithmic dependence in s can be improved by one power, at the expense of a larger factor in δ^{-1} [8]. However, this may not be best for the purposes of this paper, since in view of Theorem 3.7, δ^{-2} scales linearly in the parameter η (see next).

The following lemma estimates the constant $\sigma_{s,k}$ for matrices of the form (3.18):

Lemma 3.14. *Let $A \in \mathbb{C}^{m\times N}$ be the matrix of a bounded orthonormal system, $1 \leq s, k \leq N$ and $\sigma_{s,k}$ be as in (3.15). Then*

$$\sigma_{s,k} \leq \sqrt{\frac{K^2 sk}{m}}.$$

Proof. Fix subsets $S \subseteq \{1, \ldots, N\}$, $|S| = s$ and $T \subseteq \{1, \ldots, m\}$, $|T| = k$ and let $x \in \mathbb{C}^N$ and $c \in \mathbb{C}^m$ with $\mathrm{supp}(x) = S$ and $\mathrm{supp}(c) = T$. Then

$$\begin{aligned} |c^* Ax|^2 &= \frac{1}{\sqrt{m}} \left| \sum_{i\in T} \overline{c_i} \sum_{j \in S} \phi_j(t_i) x_j \right| \\ &\leq \frac{1}{\sqrt{m}} \max_{i=1,\ldots,m} \left| \sum_{j\in S} \phi_j(t_i) x_j \right| \sum_{i \in T} |c_i| \leq \frac{K}{\sqrt{m}} \|x\|_1 \|c\|_1 \leq \sqrt{\frac{K^2 sk}{m}} \|x\|_2 \|c\|_2. \end{aligned}$$

Hence $\|P_T A P_S\|_2 \leq \sqrt{\frac{K^2 sk}{m}}$. This now gives the result. □

With this in hand, we now deduce the following result:

Theorem 3.15. *Let $1 \leq s \leq N$, $1 \leq k \leq m$, $0 < \delta, \epsilon < 1$ and suppose that*

$$m \gtrsim \delta^{-2} \cdot K^2 \cdot s \cdot \left(\log^3(2s) \cdot \log(2N) + \log(\epsilon^{-1}) \right), \tag{3.19}$$

and

$$m \geq 8 \cdot \delta^{-2} \cdot K^2 \cdot s \cdot k.$$

Then, with probability at least $1 - \epsilon$, A has the RIP of order (s, k) for the sparse corruptions problem with constant $\delta_{s,k} \leq \delta$.

Proof. Theorem 3.13 and (3.19) imply that A has the RIP of order s with $\delta_s \leq \delta/\sqrt{2}$ with probability at least $1 - \epsilon$. Moreover, Lemma 3.14 and (3.15) imply that $\sigma_{s,k} \leq \delta/(2\sqrt{2})$. We now apply Lemma 3.11. □

Remark 3.16 This result asserts that the number of corruptions that can be tolerated is a fraction of m/s. This is inferior to the case of Gaussian random measurements, where Theorem 3.12 gives that a fraction of m corruptions are permitted. We conjecture, however, that a similar estimate can be proved for the bounded orthonormal systems case – indeed, a nonuniform recovery result of this form was proved in [21] for the case of exactly sparse coefficients x and corruptions c with random sign sequences – albeit with a substantially more sophisticated argument than the proof of Theorem 3.12. In particular, while estimates for the singular values of matrices of bounded orthonormal systems are known [38], they are more stringent than those for Gaussian random matrices. Using these estimates and arguing via the union bound (as in the proof of Theorem 3.12) unfortunately results in an estimate similar to (3.15). We also note in passing that while there exist RIP estimates for quite general matrices under the sparsity in levels model [20] (see Remark 3.9), these unfortunately do not apply to the setup of the sparse corruptions problem. We therefore leave the problem of improving (3.15) for future work.

3.4 Strategy for choosing λ

Regardless of the matrix A, our main theorems (Theorems 3.7 and 3.13) suggest an optimal strategy for choosing the parameter λ. Notice that the restricted isometry constant δ enters into the measurement condition in Theorem 3.13 as δ^{-2}. Since Theorem 3.7 requires that (3.8) holds, the measurement condition contains a factor that is at least as large as

$$1 + \left(\frac{1}{2\sqrt{2}} + \sqrt{\eta} \right)^2.$$

We wish to minimize this factor so as to reduce the measurement condition as much as possible. This can be done by minimizing η, which in turn yields the theoretically-optimal scaling

$$\lambda = \sqrt{\frac{s}{k}}. \tag{3.20}$$

Notice that this gives the value $\eta = 2$. In particular, the condition (3.8) becomes

$$\delta_{2s,2k} < \sqrt{8/33} \approx 0.492, \tag{3.21}$$

with right-hand side independent of s and k. We remark in passing that the choice (3.20) is implicitly made in [21]. However, the condition given in [21, Lem. 2.3] is $\delta_{2s,2k} < 1/18 \approx 0.056$ which is significantly more stringent than (3.20). Moreover, [21] only considers exact sparsity, whereas Theorem 3.7 also treats the case of stable recovery of inexactly sparse coefficients and corruptions.

4 Numerical experiments

We divide our numerical results into two main sections. The goal of Section 4.1 is to study the behavior of numerical algorithms in the context of the theoretical estimates presented earlier. In particular, we investigate the influence that the regularization parameter λ has on recovery properties. We confine these investigations to problems with manufactured sparsity so that systematic studies may be carried out. The lessons learned from these studies allow us to formulate and propose an iteratively reweighted alternative to the one-time optimization (2.3). Note that none of our theoretical error estimates apply to algorithms with weighted norms. However, weighted ℓ^1 schemes can provide empirically superior results, e.g., [43]. Thus, we explore weighted algorithms because their use is natural from a practical point of view, but is not in the scope of our theoretical analysis. Our simulations in this section use the SPGL1 package [36, 37].

The second collection of results, Section 4.2, focuses on more practical scenarios in scientific computing, dealing with recovery of sparse or compressible polynomial Chaos expansions of solutions to parameterized differential equations. Here we use the algorithmic lessons learned from Section 4.1 to illustrate the efficacy and fault-tolerance of our approaches on realistic problems in the presence of measurement corruptions.

4.1 Recovery of manufactured solutions with sparse corruptions

This section is primarily concerned with the generation of phase recovery diagrams for the sparse corruptions problem. In particular, our tests here are not necessarily motivated by sensing matrices and corruptions from function approximation, but instead are designed to understand behavior of the algorithms. The following standard experiment for accomplishing this is carried out: We fix the number of measurements m and the dictionary size N, and we vary the signal sparsity s and the number of measurement corruptions k. For each s and k we generate an s-sparse signal x, and for a given model of a measurement matrix A, we generate m measurements y from the signal x, and subsequently corrupt (highly pollute) these measurements with a k-sparse vector c, whose non-zero entries are CZ, where Z is a random draw from a certain probability distribution and $C > 0$ is a scaling constant. In this test, Z is a standard normal random variable and $C = 1$.

We then run the recovery algorithm (2.3) for a given value of λ, producing a recovered signal $\widehat{x}$ and measurement corruption vector $\widehat{c}$. We define the recovery as successful if $\|x - \widehat{x}\|^2 + \|c - \widehat{c}\|^2 < \epsilon_{\text{tol}}$. In this test, we set the success tolerance to be $\epsilon_{\text{tol}} = 10^{-4}$.

In the test above, the generation of x, and of y, and of c, are statistically independent[2]. For each s and k, the above procedure is run $T \in \mathbb{N}$ times with independent draws, and an empirical estimate of the probability of "success" is computed. In the phase transitions plots below, we use $T = 10$ simulations.

The phase transitions color each pixel, corresponding to a particular value of s and k, according to the empirical success probability. The phase transition axes are s/m and k/m, and thus each ranges in the interval $[0, 1]$, but we truncate to $[0, 0.5]$ in our plots because this region is sufficient to illustrate behavior. We consider the following two models of measurement matrix A:

- Model 1: a Gaussian random matrix
- Model 2: a randomly-subsampled Discrete Fourier Transform (DFT) matrix

Note that Model 2 is an example of a bounded orthonormal system. We compare several different choices of λ for each model.

4.1.1 Phase transition plots for fixed λ

Figures 1 and 2 display the results for models 1 and 2 described above, respectively. Each figure shows an array of plots; the columns correspond to differing values of m, increasing from left to right; the rows correspond to differing values of λ, increasing from top to bottom, except the last two rows, which show the "optimal" value of $\lambda = \sqrt{s/k}$ suggested by the theory, and the iterative reweighting procedure described in the next section.

[2]Measurement corruptions are generated as iid standard normal random variables, and support indices in a sparse vector are generated using the uniform probability law (draws without replacement) on the index set.

Comparing the results for $\lambda = \sqrt{s/k}$ (row 5 in the plots) with the other plots with λ fixed, we see that $\lambda = \sqrt{s/k}$ does not behave optimally in practice, even though this is suggested by our theory. Indeed, further experimentation reveals that the behavior of these transition plots changes notably when m is varied. However, the following observations are consistent across all our runs:

- When there are few corruptions relative to the signal sparsity ($k \ll s$), larger values of λ tend to perform better. This general trend is consistent with the theory from previous sections: Our recovery results are stated in terms of a quantity η defined in (3.4), and when $k \ll s$, we require large λ to make η small.
- When there are many corruptions relative to signal sparsity ($k \sim s$), smaller values of λ tend to perform better. Again, this is consistent with the theory in terms of the parameter η.

4.1.2 Iteratively reweighted ℓ^1 minimization

The results from the previous section show that our *a priori* postulated optimal values of λ are not optimal in practice; this suggests that an adaptive learning of λ may produce better results. See, for example, [7]. This section introduces an iteratively reweighted ℓ^1 optimization procedure that effects this learning of λ.

We compute minimizers $\hat{x}$ and $\hat{c}$ using an initial value of λ. We then update λ based on $\hat{x}$ and $\hat{c}$, and then recompute minimizers with the new λ. Such an approach not only allows for a single parameter λ to be updated, it also permits individual (i.e. non-equal) weights to be used for term in the regularization functional. This aims to enhance recovery performance by both iteratively estimating an optimal weighting λ between the coefficients and corruptions term, and iteratively estimating the support sets of x and c.

We outline the procedure below:

- Step 1. Set $r = 1$, $\mu_i = 1$ for $i = 1, \dots, N$, and $\lambda_j = 1$ for $j = 1, \dots, m$. Prescribe noise tolerance ϵ and a small positive number $\eta > 0$.
- Step 2. Compute the solution $(\widehat{x}, \widehat{c})$ to

$$\min_{z \in \mathbb{C}^N, d \in \mathbb{C}^m} \|z\|_{1,\mu} + \|d\|_{1,\lambda} \text{ subject to } \|Az + d - y\|_2 \leq \epsilon,$$

  where $\|z\|_{1,\mu} = \sum_{i=1}^N \mu_i |z_i|$ and $\|d\|_{1,\lambda} = \sum_{j=1}^m \lambda_j |d_j|$.
- Step 3. Update μ and λ as follows:

$$\mu_i = \frac{1}{\eta + |\hat{x}_i|}, \qquad \lambda_i = \frac{1}{\eta + |\hat{c}_i|}. \tag{4.1}$$

- Step 4. If $r < r_{\max}$, set $r = r + 1$ and go back to step 2, otherwise stop.

Numerical results in the bottom row of plots in Figures 1 and 2 show this approach (implemented with $r_{\max} = 10$ iterations) significantly improves the recovery over a fixed choice of λ. We therefore use this iteratively reweighted ℓ^1 approach for optimization for all our simulations in the next section.

4.1.3 Large corruption values

This section is devoted to understanding the behavior of our algorithm with respect to the magnitude of the corruptions.

We run the same experiment as outlined at the beginning of Section 4.1 on Model 2 (the measurement matrix is a subsampled DFT matrix) using the iteratively reweighted algorithm outlined in Section 4.1.2. For this test, we vary C between 1 and 10^6, and choose the random variable Z defining the corruptions as a standard Cauchy random variable.[3]

[3]The point of generating from a Cauchy distribution is to show that measurement corruption by heavy-tailed distributions does not adversely affect the algorithm's results.

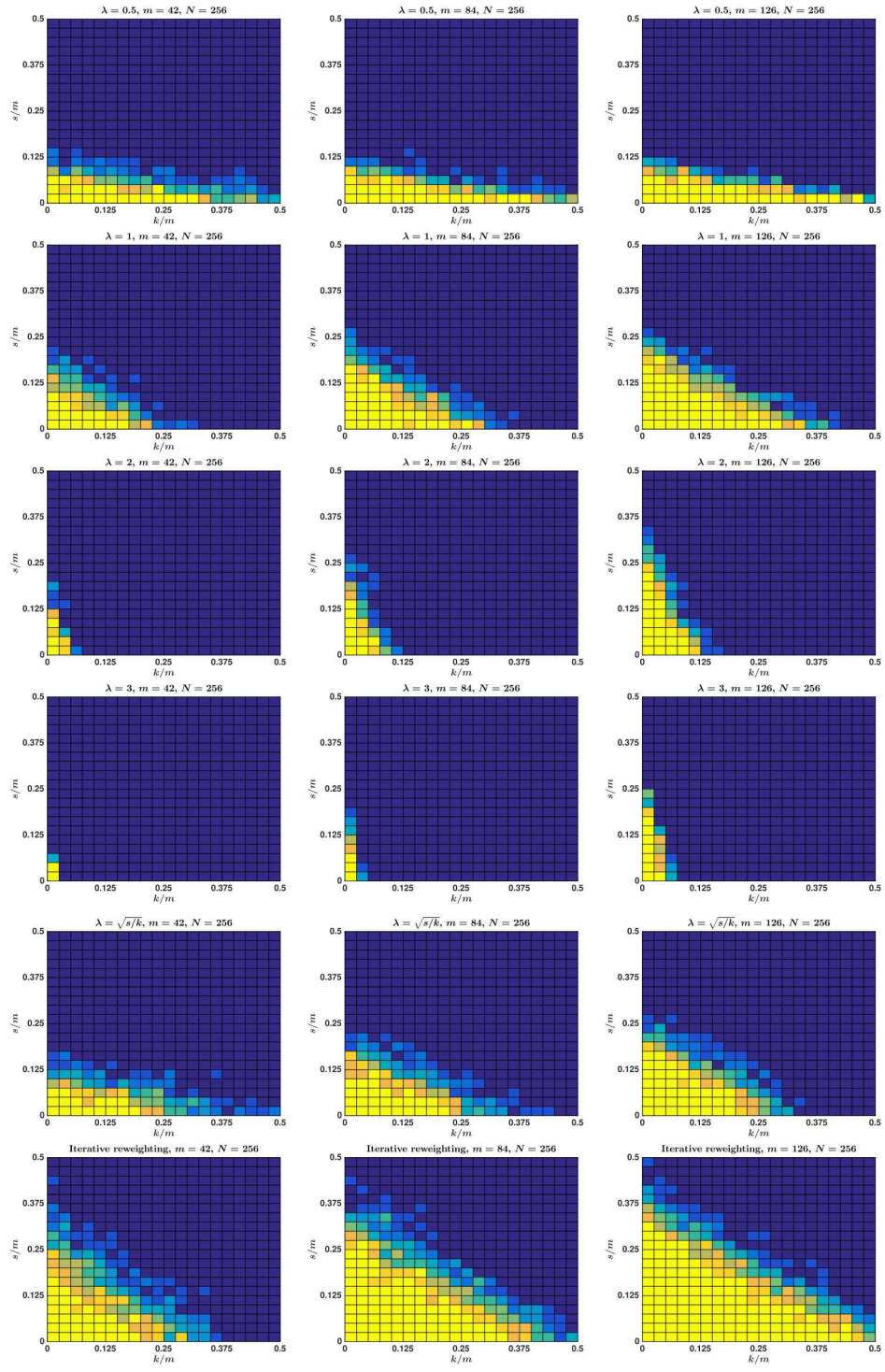

Figure 1: Phase transition for model 1 with fixed $N = 256$, varying m and λ. Each column represents varying values of m: from left to right, $m = 42$, $m = 84$, and $m = 126$. Each row represents different values of λ: rows 1-4 correspond to $\lambda = 0.5, 1, 2, 3$, respectively. Row 5 uses the value $\lambda = \sqrt{s/k}$ that is suggested as optimal by the theory. Row 6 shows recovery using the iteratively reweighted ℓ^1 algorithm. Each pixel is colored according to its probability of a successful signal recovery for $T = 10$ trials based on repeated random draws of x and c; yellow is probability 1, blue is probability 0.

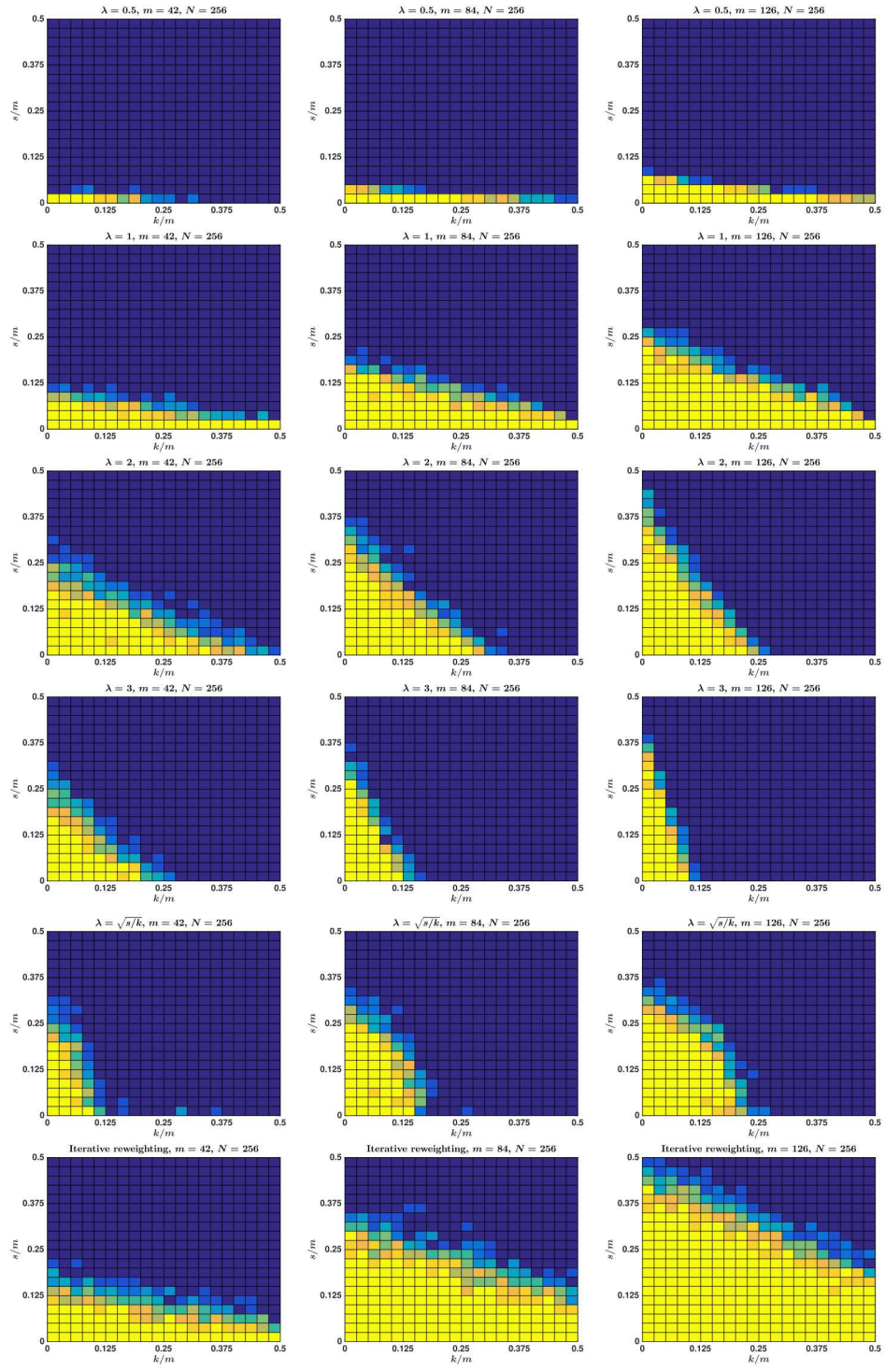

Figure 2: Phase transition for model 2 with fixed $N = 256$, varying m and λ. Each column represents varying values of m: from left to right, $m = 42$, $m = 84$, and $m = 126$. Each row represents different values of λ: rows 1-4 correspond to $\lambda = 0.5, 1, 2, 3$, respectively. Row 5 uses the value $\lambda = \sqrt{s/k}$ that is suggested as optimal by the theory. Row 6 shows recovery using the iteratively reweighted ℓ^1 algorithm. Each pixel is colored according to its probability of a successful signal recovery for $T = 10$ trials based on repeated random draws of x and c; yellow is probability 1, blue is probability 0.

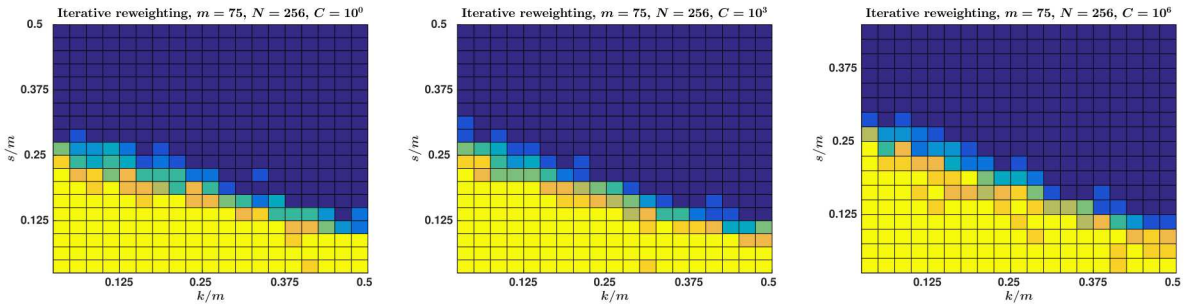

Figure 3: Phase transition for model 2 with fixed $N = 256$ and $m = 75$, varying the corruptions magnitude C. (Left: $C = 1$. Middle: $C = 10^3$. Right: $C = 10^6$.) Each transition plot uses the iteratively reweighted algorithm outlined in Sections 4.1.2 with the augmentations described in Section 4.1.3. The recovery property of the corruptions algorithm is relatively agnostic to magnitude of the corruptions.

A straightforward application of the iteratively reweighted algorithm in Section 4.1.2 when C is very large produces suboptimal results. The reason for this is the scale differential between x and c, so that the algorithm heavily favors recovery of the corruptions and devotes little effort to recovering the signal. To overcome this limitation, we leverage a significant advantage of our algorithm: Corruption indices and values in the measurement vector are identified. This allows us to formulate a slight modification of the algorithm in Section 4.1.2:

1. Run the algorithm from Section 4.1.2, generating computed solutions $\widehat{x}$ and $\widehat{c}$.
2. If $\|\widehat{c}\| < C_{\max}\|y - \widehat{c}\|$, then return the solutions $\widehat{x}$ and $\widehat{c}$.
3. If instead $\|\widehat{c}\| \geq C_{\max}\|\widehat{y} - \widehat{c}\|$, then define a support set for the vector $\widehat{c}$ as
$$S = \{j = 1, \ldots, m \;\mid\; \widehat{c}_j \geq \tau \, \|\widehat{c}\|\},$$
and let $\widehat{c}_S$ equal to $\widehat{c}$ on S and zero otherwise.
4. Remove the large corruptions from the measurements and resolve with the measurements $\widetilde{y} \leftarrow y - \widehat{c}_S$. This yields a new solution pair $\widetilde{x}$ and $\widetilde{c}$. Return $x = \widetilde{x}$ and $c = \widehat{c}_S + \widetilde{c}$.

This procedure uses the algorithm to identify and remove highly corrupted measurements, and then uses another instance of the algorithm to accurately compute the signal. We use the procedure above with the choices $C_{\max} = 10$ and $\tau = \frac{1}{5\sqrt{m}}$.

We can now generate a phase transition plot for a fixed value of C. Figure 3 shows the transition plots for values $C = 1, 10^3$, and 10^6. We see that the algorithm detects and removes corruptions just as well when $C = 1$ as when $C = 10^6$.

Remark 4.1 The iteratively reweighted procedure in (4.1) updates weights for both the corruptions (λ_i) and the signal (μ_i). Since our focus here is recovery of the corruptions, one may wonder which set of weights is more influential. We have conducted tests in this direction by performing an experiment parallel to the results in Figure 3, where we iteratively update λ_i according to (4.1), but keep μ_i fixed at unity for all i. Our results, shown in Figure 4, indicate that fixing the weights μ_i results in significant deterioration of the algorithm's performance when $C = 1$. However, it results in notable improvement of the algorithm when $C = 10^3$ or $C = 10^6$. In the context of soft faults, the $C = 1$ behavior of the algorithm is more relevant since when $C \geq 10^3$ it is likely that the corruptions can be easily identified and removed by other means. In this small-C context, allowing both sets of weights μ_i and λ_i to vary appears to be beneficial. On the other hand, the deterioration of the algorithm for very large C is an interesting phenomenon whose investigation we leave for future work.

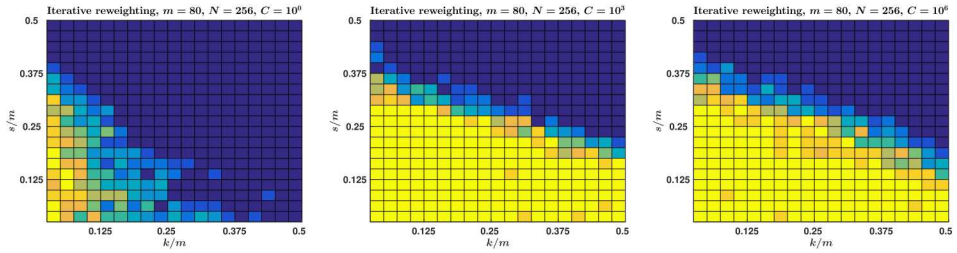

Figure 4: Diagram complementary to Figure 3, here using fixed signal weights $\mu_i = 1$ but varying the corruptions weights λ_i in the iteratively reweighted algorithm described in Sections 4.1.2 and 4.1.3. Phase transition for model 2 with fixed $N = 256$ and $m = 80$, varying the corruptions magnitude C. (Left: $C = 1$. Middle: $C = 10^3$. Right: $C = 10^6$.) The results indicate empirical superiority of the algorithm in Section 4.1.2 that allows both μ_i and λ_i to vary, compared with fixing μ_i.

4.2 Recovery of compressible polynomial Chaos expansions

In this section we test our algorithm on more realistic problems in UQ: sparse recovery of multivariate polynomial Chaos expansion coefficients with corrupted measurements. Polynomial chaos expansions (PCE) [40, 14] have become a popular means of quantifying parametric uncertainty in expensive computer simulations. To formulate our problem using our earlier notation, let $f(\xi)$ denote a scalar-valued response of a model (e.g., a differential equation) where $\xi \in \mathbb{R}^d$ is a random parameter appearing in the model. The dependence of y on ξ thus encodes uncertainty in the response. We are interested in building the approximation $\xi \mapsto \sum_{n=1}^{N} x_n \phi_n(\xi)$, where $\{\phi_n\}_{n=1}^{N}$ are computable orthonormal polynomials constructed from the probability density of the random vector ξ, and we wish to compute the unknown coefficients x_n. In a CS recovery procedure, we construct m samples $\{\xi_j\}_{j=1}^{m}$ of the random vector ξ, collect the measurements $y_j = f(\xi_j)$, and then attempt to find a sparse coefficient vector x minimizing $\|y - Ax\|$, where A is the measurement matrix with entries $(A)_{j,n} = \phi_n(\xi_j)$. The underlying assumption is that $\xi \mapsto f(\xi)$ is expensive to evaluate so that m should be as small as possible. To focus our study on the corruptions problem, we consider the case where the vector y can have a sparse number of entries that are polluted by large-magnitude errors.

The models $f(\xi)$ we consider here reflect the types of large scale models that are susceptible to soft failures. However, these test models can be evaluated repeatedly with almost zero probability of corruptions. Therefore, to simulate the effect of soft failures we randomly generate soft faults according to the corruptions model from Section 4.1. After constructing components of y as $f(\xi)$, we pollute k of these entries as described at the beginning of Section 4.1. In our tests below we fix a value $r := k/m$, the ratio of corrupted measurements.

4.2.1 Genz test functions

We compare the algorithm presented in this paper against a classical ℓ^1 minimization approach in the presence of measurement corruptions for the purposes of computing compressible PCE expansion coefficients of a function. A classical ℓ^1 minimization algorithm sets the corruptions vector $d = 0$ in (2.3) and minimizes over all $x \in \mathbb{R}^N$.

Our function $f(\xi)$ will be one of the multidimensional test functions used by Genz [13]. For $\xi \in \mathbb{R}^d$, $d \in \mathbb{N}$, we investigate computing expansion coefficients for the following two functions on the hypercube

$[-1,1]^d$:

$$f(\xi) = \exp\left[-\frac{2}{\sqrt{d}}\sum_{j=1}^{d}(\xi_j - w_j)^2\right], \qquad w_j = \frac{(-1)^j}{j+1}, \quad (\text{"Gaussian"})$$

$$f(\xi) = \prod_{j=1}^{d}\frac{d/4}{d/4 + (\xi - w_j)^2}, \qquad w_j = \frac{(-1)^j}{j+1}, \quad (\text{"Product Peak"})$$

We use $d = 4$ and $d = 10$ in our tests, with the dictionary elements ϕ_n given by tensor-product Chebyshev polynomials of total degree 10 and 4, respectively, over $[-1,1]^d$. We set the corruptions ratio to the value $r = 0.1$ uniformly over all tests, and vary the corruptions magnitude C. After computing a coefficient vector x solving either a classical ℓ^1 problem or (2.3), we compute a discrete ℓ^2 error metric defined by

$$\sqrt{\frac{1}{Q}\sum_{q=1}^{Q}(f_N(\tau_q) - f(\tau_q))^2}, \qquad f_N(\xi) \coloneqq \sum_{n=1}^{N} x_n \phi_n(\xi)$$

where $Q = 10^3$ for each test, and τ_q are iid samples drawn from the product Chebyshev distribution over $[-1,1]^d$.

Figure 5 shows the result of this test. (See the figure caption for additional details of the test.) The results indicate that when corruptions are present, a standard ℓ^1 minimization algorithm suffers severe degradation of the quality of the computed expansion coefficients. However, the corruptions algorithm of this paper is able to compute accurate coefficients in the presence of corruptions, whether they have large or small magnitude.

This example shows that there may be a penalty for using our algorithm when no corruptions are present. This is mostly easily noticed in the product peak example with no corruptions ($C = 0$): The corruptions algorithm of this paper computes a PCE that is less accurate than the result using a standard ℓ^1 minimization approach. (Compare the black lines in row 3 of Figure 5.)

4.2.2 Damped Harmonic Oscillator

In this section we investigate the fault-tolerance of our algorithm for recovery of PCE coefficients in a damped linear oscillator subject to external forcing with six unknown parameters. The model is

$$\begin{aligned} &\frac{d^2u}{dt^2}(t,\xi) + \gamma\frac{du}{dt} + ku = g\cos(\omega t), \\ &u(0,\xi) = u_0(\xi), \quad \dot{u}(0,\xi) = u_1(\xi), \end{aligned} \tag{4.2}$$

where we assume the damping coefficient γ, spring constant k, forcing amplitude g and frequency ω, and the initial conditions u_0 and u_1 are all uncertain, defining components of a 6-dimensional random vector ξ. We solve (4.2) analytically to circumvent the impact of discretization errors in our study.

Defining $\xi = (\gamma, k, g, \omega, u_0, u_1)$, we restrict the components $\xi^{(j)}$ of ξ to the following ranges:

$$\begin{aligned} &\xi^{(1)} \in [0.08, 0.12], && \xi^{(2)} \in [0.03, 0.04], && \xi^{(3)} \in [0.08, 0.12], \\ &\xi^{(4)} \in [0.8, 1.2], && \xi^{(5)} \in [0.45, 0.55], && \xi^{(6)} \in [-0.05, 0.05]. \end{aligned}$$

We define $I_\xi \in \mathbb{R}^6$ to be the range of ξ defined by the product of these intervals. For any parameter realization in I_ξ the harmonic oscillator is underdamped. In the following, we choose our quantity of interest as $f(\xi) = u(20, \xi)$. We set the corruptions magnitude C as the mean of the function, i.e. $C = \mathbb{E}_\xi[f]$.

Figure 6 compares, as a function of the number of measurements, the error in classical ℓ^1 recovery for uncorrupted sparse recovery versus the iteratively reweighted version of the sparse corruptions ℓ^1 optimization proposed in Section 4.1.2. The results show that the sparse corruptions optimization notably outperforms standard ℓ^1 minimization when corruptions are present, and is competitive without corruptions.

In Figure 7 we run the iteratively reweighted sparse corruptions optimization but vary the corruptions rate r, and the corruptions magnitude C. The left-hand plot shows predictable behavior: increasing

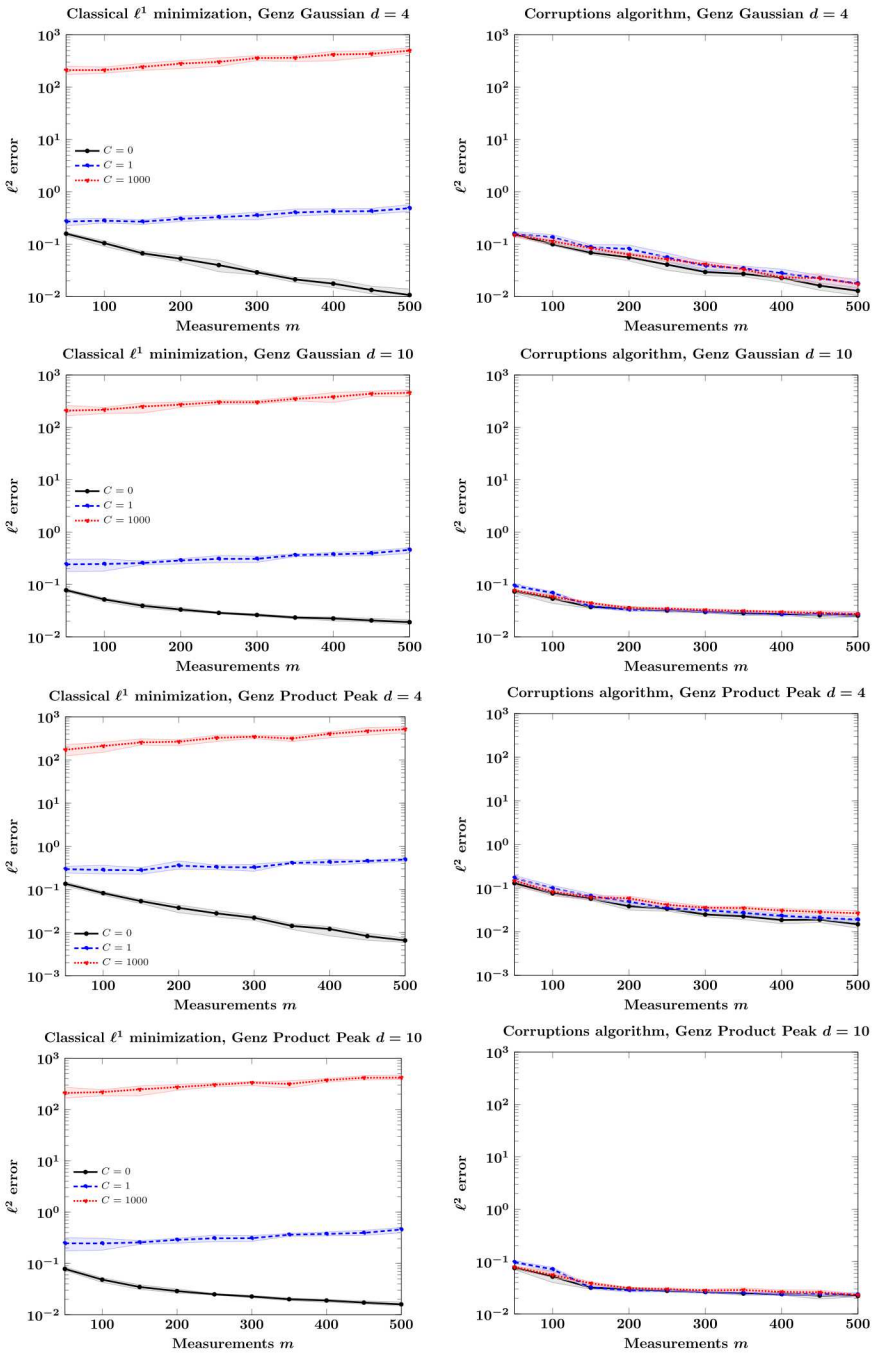

Figure 5: Approximation of sparse representations for Genz test functions in the presence of measurement corruptions. Left: classical ℓ^1 minimization. Right: The corruptions algorithm of this paper. The top two rows use a Genz Gaussian test function ($d = 4$ and $d = 10$), the bottom two rows use a Genz Product Peak test function ($d = 4$ and $d = 10$). 10% of the measurements are corrupted in each test ($r = 0.1$), with varying values of the corruptions magnitude C. Results over a size $T = 10$ ensemble are shown, with the mean error plotted with a solid curve, and shaded regions around the mean demarcated by the 20% and 80% quantiles.

corruptions has deleterious effects on the error in recovery, but notably the algorithm is reasonably stable for increasing r. The right-hand plot shows that the algorithm is relatively insensitive to the magnitude of the corruptions.

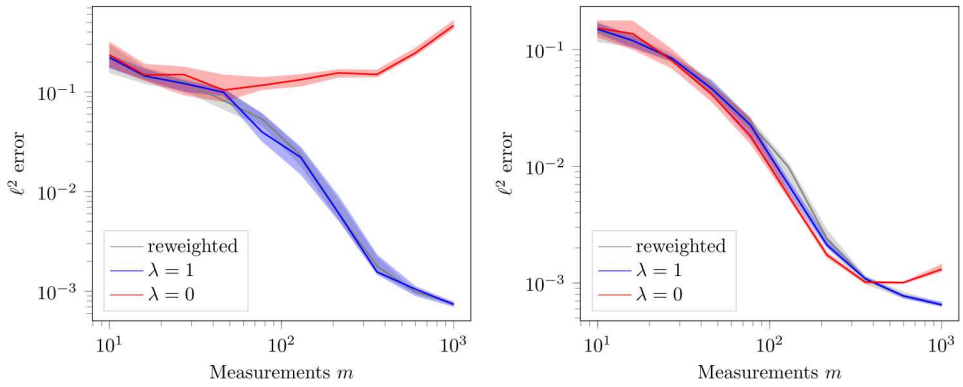

Figure 6: Comparison of iteratively reweighted ℓ_1-minimization with classical ℓ_1-minimization ($\lambda = 0$) when constructing a PCE of the $d = 6$ harmonic oscillator in the presence of (left) corrupted data with $r = 0.1$ and $C = 1$ and (right) no failures.

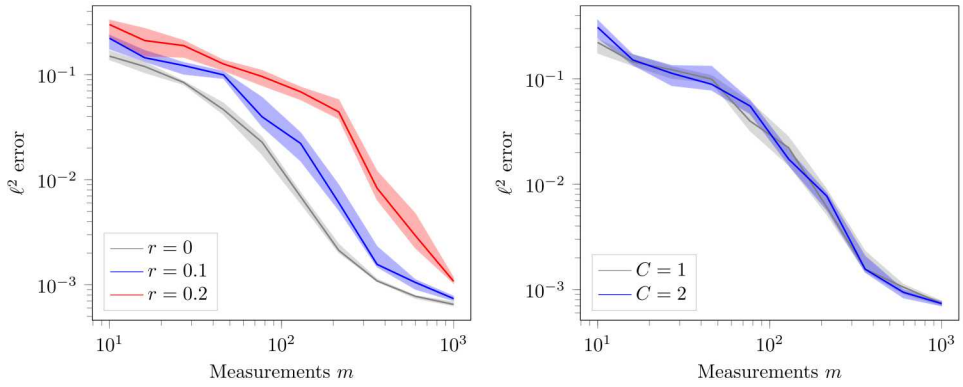

Figure 7: Effect of the corruption rate r (left) and magnitude C of corruption errors (right) on the PCE of the $d = 6$ harmonic oscillator constructed in the presence of failures using ℓ_1-minimization with various choices of λ. To generate the left and rights plot we set $C = 1$ and $r = 0.1$, respectively.

5 Summary and conclusion

We have developed novel theoretical guarantees and algorithms for recovery of sparse or compressible signals where measurements have been polluted by high-magnitude corruptions. Our results are uniform theoretical recovery estimates for general linear systems where the measurement matrix satisfies a corruptions-based RIP-like condition.

We have refined an existing regularized ℓ^1 minimization algorithm into an iteratively reweighted ℓ^1 minimization algorithm that shows superior performance for the examples that we have investigated. An application of these examples to recovery of polynomial Chaos expansions from model UQ problems illustrates that our algorithms are resistant to highly-corrupted measurement data that may result from hardware or software faults in modern large-scale parallel computing paradigms.

Empirical tests suggest that refinements of our algorithm is relatively stable with respect to the magnitude of the corruptions, but our theory is not applicable to these algorithmic refinements and some

observed behavior (e.g., Remark 4.1) remains theoretically unexplained, which can be the subject of future explorations.

Acknowledgments

B. Adcock thanks Simone Brugiapaglia and Xiaodong Li for helpful discussions. The authors acknowledge an anonymous referee whose report led to the investigations outlined in Remark 4.1.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

References

[1] B. Adcock. Infinite-dimensional compressed sensing and function interpolation. *Found. Comput. Math. (to appear)*, 2017.

[2] B. Adcock. Infinite-dimensional ℓ^1 minimization and function approximation from pointwise data. *Constr. Approx. (to appear)*, 2017.

[3] B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. *Forum of Math., Sigma (to appear)*, 2017.

[4] B. Adcock, A. C. Hansen, and B. Roman. The quest for optimal sampling: computationally efficient, structure-exploiting measurements for compressed sensing. In *Compressed Sensing and Its Applications*. Springer, 2015.

[5] A. Bastounis and A. C. Hansen. On the absence of the RIP in real-world applications of compressed sensing and the RIP in levels. *arXiv:1411.4449*, 2014.

[6] P. G. Bridges, K. B. Ferreira, M. A. Heroux, and M. Hoemmen. Fault-tolerant linear solvers via selective reliability. *arXiv:1206.1390 [cs, math]*, June 2012. arXiv: 1206.1390.

[7] E. J. Candès, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.*, 14(5):877–905, 2008.

[8] A. Chkifa, N. Dexter, H. Tran, and C. Webster. Polynomial approximation via compressed sensing of high-dimensional functions on lower sets. Technical Report ORNL/TM-2015/497, Oak Ridge National Laboratory, 2015.

[9] I.-Y. Chun and B. Adcock. Compressed sensing and parallel acquisition. *arXiv:1601.06214*, 2016.

[10] A. Doostan and H. Owhadi. A non-adapted sparse approximation of PDEs with stochastic inputs. *Journal of Computational Physics*, 230(8):3015–3034, Apr. 2011.

[11] D. Dorsch and H. Rauhut. Refined analysis of sparse mimo radar. *J. Fourier Anal. Appl.*, pages 1–45, 2016.

[12] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhauser, 2013.

[13] A. Genz. Testing multidimensional integration routines. In *Proc. of international conference on Tools, methods and languages for scientific and engineering computation*, pages 81–94, New York, NY, USA, 1984. Elsevier North-Holland, Inc.

[14] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag New York, Inc., 1991.

[15] L. Guo, A. Narayan, T. Zhou, and Y. Chen. Stochastic collocation methods via L_1 minimization using randomized quadratures. *SIAM Journal on Scientific Computing (accepted) arXiv:1602.00995 [math]*, Feb. 2016. arXiv: 1602.00995 [math.NA].

[16] J. Hampton and A. Doostan. Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies. *Journal of Computational Physics*, 280:363–386, Jan. 2015.

[17] J. D. Jakeman, M. S. Eldred, and K. Sargsyan. Enhancing ℓ_1-minimization estimates of polynomial chaos expansions using basis selection. *arXiv:1407.8093*, 2014.

[18] J. D. Jakeman, A. Narayan, and T. Zhou. A generalized sampling and preconditioning scheme for sparse approximation of polynomial chaos expansions. *SIAM Journal on Scientific Computing (to appear), arXiv:1602.06879 [math]*, Feb. 2016. arXiv: 1602.06879 [math.NA].

[19] J. Laska, M. A. Davenport, and R. G. Baraniuk. Exact signal recovery from sparsely corrupted measurements through the pursuit of justice. In *Asilomar Conference on Signals Systems and Computers*, 2009.

[20] C. Li and B. Adcock. Compressed sensing with local structure: uniform recovery guarantees for the sparsity in levels class. *arXiv:1601.01988*, 2016.

[21] X. Li. Compressed sensing and matrix completion with a constant proportion of corruptions. *Constr. Approx.*, 37:73–99, 2013.

[22] A. Narayan and T. Zhou. Stochastic collocation on unstructured multivariate meshes. *Commun. Comput. Phys.*, 18(1):1–36, 2015.

[23] T. Nguyen and T. D. Tran. Exact recoverability from dense corrupted observations via ℓ_1-minimization. *IEEE Trans. Inform. Theory*, 59(4):2017–2035, 2013.

[24] S. Pauli. *Fault Tolerance in Multilevel Monte Carlo Methods*. PhD thesis, ETH Zurich, Switzerland, 2014.

[25] S. Pauli, P. Arbenz, and C. Schwab. Intrinsic fault tolerance of multilevel Monte Carlo methods. *Journal of Parallel and Distributed Computing*, 84:24–36, Oct. 2015.

[26] S. Pauli, M. Kohler, and P. Arbenz. A fault tolerant implementation of multi-level Monte Carlo methods. In M. Bader, editor, *Parallel Computing: Accelerating Computational Science and Engineering (CSE)*, pages 471–480. IOS Press, 2014. doi:10.3233/978-1-61499-381-0-471.

[27] J. Peng, J. Hampton, and A. Doostan. A weighted ℓ_1-minimization approach for sparse polynomial chaos expansions. *Journal of Computational Physics*, 267:92–111, June 2014.

[28] H. Rauhut and R. Ward. Sparse Legendre expansions via l1-minimization. *J. Approx. Theory*, 164(5):517–533, 2012.

[29] H. Rauhut and R. Ward. Interpolation via weighted ℓ_1 minimization. *Appl. Comput. Harmon. Anal.*, 40(2):321–351, 2016.

[30] B. Roman, A. C. Hansen, and B. Adcock. On asymptotic structure in compressed sensing. *arXiv:1406.4178*, 2014.

[31] Y. Shin and D. Xiu. Correcting Data Corruption Errors for Multivariate Function Approximation. *SIAM Journal on Scientific Computing*, 38(4):A2492–A2511, Jan. 2016.

[32] L. Stankovic, S. Stankovic, and M. Amin. Missing samples analysis in signals for applications to L-estimation and compressive sensing. *Signal Processing*, 94:401–408, Jan. 2014.

[33] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei. Recovery of sparsely corrupted signals. *IEEE Trans. Inform. Theory*, 58(5):3115–3130, 2012.

[34] D. Su. Compressed sensing with corrupted Fourier measurements. *arXiv:1607.04926*, 2016.

[35] D. Su. Data recovery from corrupted observations via l1 minimization. *arXiv:1601.06011*, 2016.

[36] E. van den Berg and M. P. Friedlander. SPGL1: A solver for large-scale sparse reconstruction, June 2007. http://www.cs.ubc.ca/labs/scl/spgl1.

[37] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.

[38] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*, chapter 5, pages 210–268. Cambridge University Press, Cambridge, U.K., 2012.

[39] J. Wright and Y. Ma. Dense correction via ℓ_1-minimization. *IEEE Trans. Inform. Theory*, 56(7):3540–3560, 2010.

[40] D. Xiu and G. E. Karniadakis. The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, Jan. 2002.

[41] L. Yan, L. Guo, and D. Xiu. Stochastic collocation algorithms using ℓ_1-minimization. *International Journal for Uncertainty Quantification*, 2(3):279–293, 2012.

[42] J. Yang and Y. Zhang. Alternating Direction Algorithms for ℓ_1-Problems in Compressive Sensing. *SIAM Journal on Scientific Computing*, 33(1):250–278, Jan. 2011.

[43] X. Yang and G. E. Karniadakis. Reweighted ℓ_1 minimization method for stochastic elliptic differential equations. *Journal of Computational Physics*, 248:87–108, Sept. 2013.

[44] P. Yin, Y. Lou, Q. He, and J. Xin. Minimization of ℓ_{1-2} for Compressed Sensing. *SIAM Journal on Scientific Computing*, 37(1):A536–A563, Jan. 2015.